UNDERSTANDING STRUCTURE AND DYNAMICS OF PTEN AND ITS POSSIBLE GENOTYPE-PHENOTYPE CORRELATIONS IN ENDOMETRIOSIS AND CANCER

A Dissertation Presented to

the Faculty of the Department of Biology and Biochemistry

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Iris Nira Smith

December 2016

Dedication

This dissertation is dedicated to my husband Jeff as well as my parents Flavio and Mary Jane Nira, who have supported and encouraged me to follow my passion for science and to trust in God's sovereignty in spite of the health challenges that I've endured as a result of being afflicted with stage IV endometriosis.

UNDERSTANDING STRUCTURE AND DYNAMICS OF PTEN AND ITS POSSIBLE GENOTYPE-PHENOTYPE CORRELATIONS IN ENDOMETRIOSIS AND CANCER

_____
Iris Nira Smith

APPROVED:

_____
Dr. James M. Briggs, Chairman

_____
Dr. Michael J. Heard

_____
Dr. Zhang Weihua

_____
Dr. Xiaolian Gao

_____
Dr. Dan E. Wells, Dean
College of Natural Sciences and Mathematics

Acknowledgements

Where do I begin? There are so many individuals that I'd like to express my sincerest gratitude to for all of their support and encouragement. However, I must first give praise and thanks to God for granting me the opportunity and providing the financial means to pursue my passion in research. Whatever has come to pass from having endometriosis, I know it has been because of His will, my good and His glory – while in the crux of my pain at times I was not able to see His hand or hear His voice, I know He was guiding me. Looking back over the 20 years that I have been plagued with this disease, complications, and unexpected diagnoses of other afflictions that ensued, I can humbly say to Him, "just and true are Your ways!" For out of the pains of endometriosis, came my passion for research to obtain insights into the mechanism of this terrible disease. In the midst of my pain, He has brought individuals into my life to help me grow in wisdom, peace, and joy – shaping me into the person I am today. I would not change any course of events in my journey with my health or scientific research, for it has taught me that God is sovereign and in total control of all things: past, present, and future - before I had only heard of Him (Job 42:5), but know I can attest of His love for me (Jeremiah 31:3) and that all things happen in His perfect timing, for His purpose, and for my good (Romans 8:28).

I cannot express enough how grateful I am to my mentor, **Dr. James M. Briggs** for his guidance (both professionally and personally) and for investing and believing in me as a student so early on. I still remember the day I approached him in 2009 as an undergraduate asking him to petition on my behalf to enroll in a graduate course, as well as requesting to intern in his lab – thank you for taking a chance on me. Through our discussions you have challenged me to think critically about my research, in gaining a practical understanding of the research field. Thank you for your patience and support when I was out for countless doctors appointments and surgical procedures and for checking in on me during my recoveries.

To my committee members: **Drs. Xiaolian Gao**, **Michael J. Heard and Zhang Weihua** – the completion of this dissertation is a testament to your guidance, feedback, and valuable insight throughout the duration of my research and in bringing me to this point of my career. I am also grateful to each of you for your compassion and understanding of my health issues. **Dr. Weihua**, I was further empowered as a scientist after taking your Cancer Biochemistry class. Thank you for further solidifying my passion for research – your zeal has shown me the true impact we can make as cancer researchers. **Dr. Gao**, thank you for believing in me when I didn't believe in myself. Your compassion and empathy strengthened me every step of the way – I look forward to an ongoing mentor relationship and future collaboration in the fight against endometriosis and cancer. **Dr. Heard**, words can't fully explain the impact you have made on my life. You have shown me what true compassion is about - not only are you my physician, but a wonderful friend to Jeff and I. You have been my cheerleader in good times and bad – you are truly a blessing!

I'd like to thank past and present members of the Dr. Briggs lab who have supported me over the years when I felt like giving up. Collectively you have taught me that, *"Being genetically related doesn't make you family. Love, support, trust, sacrifice, honesty, protection, acceptance, security, gratitude, respect and loyalty is what makes you family."* **Dr. Jerry Ebalunode**, thank you for your assistance and patience when I needed to troubleshoot hardware/software and cluster issues. Most importantly, thank you for your prayers of encouragement (Joshua 1:9) and spiritual guidance (Luke 12:48) when I needed it most. **Dr. Prema Latha Mallipeddi,** thank you for your mentorship and prayers – as well as your guidance and support when I needed to make the hardest decision of my life. Thank you to the lab "oracle", **Dr. Tsai-Wei Shen**, your kind attitude and smile still lingers in my mind to this day when I think about how helpful you were with my understanding of UHBD modules. **Dr. Marc Chardenoff**, thank you for your support and advice and for always making me laugh – your laughter and jokes still play like a record in my mind. **Dr.**

John W. Craft, what can I say – you have been so critical to my success in the lab, your support, prayers, tears, and encouragement mean more than you can ever imagine and will forever be cherished in my heart. You have lived out and taught me the true meaning of Hebrews 6:10. **Dr. Guedmiller Souza de Oliveira**, there aren't enough words to express how thankful I am to God for bringing you into the lab – having you live with Jeff, Kya, Koda and I was such a blessing. Thank you for your prayers, support and for inviting me to Brazil to share my research. Spending the weekend with your family and enjoying the Sabbath was a blessing – thank you for watching over me, you will always be my brother. **Ben Skidmore**, thank you for taking care of me when I was in such dire pain and crawling on the lab floor. I'll never forget how fast Amir drove to get me pain meds. Most importantly, thank you for your support and guidance in everything – I'm going to miss our talks, you are wise beyond your years! My dear **Yanyun Liu**, your support, prayers and encouragement have carried me in more ways than you can imagine. Thank you for reminding me to take care of my body when I was in pain and for the memories and laughter that I will forever cherish from our conversations. My *behan*, **Khushboo Singh**, you have been my support and at times my "life jacket". You have helped me keep my head above water as I tried to stay afloat from the pain. Reminding me of God's power and sovereignty is what sustained me - you will forever be my dear sister. **Sladja Maric**, my kindred spirit and dear friend, it has been so great to have you as part of our lab family. Your prayers have been such a huge blessing – thank you for your encouragement when I needed it most. **Alison Vicary**, thank you so much for your prayers, it was truly a pleasure to mentor you - watching you grow in knowledge has further solidified my passion for mentoring. **Deepika Kumar** – honorary lab member, thank you for your friendship, encouragement, and compassion – you have been an incredible blessing. **Jeff Mindrebo, Amir Ali**, and **Lindsey Brier,** although you have journeyed on to greater things; y'all will forever hold a place in my heart. I will miss our talks and the countless times we laughed together – I will never hear the word "magenta" again without laughing out loud. To all the other

high school and undergraduate students, **Tangen Tran, Maria Williams, Cameron Lee, Masa and Marwa Kharboutli, Shivali Narang, Nar Murali, Juliette Cao, Julia C. Kirsten, Melanie Lopez, and Sara Abacellar** that I have been afforded the wonderful opportunity to mentor, thank you for your enthusiasm in conducting research in the lab – each of you has helped me realize and grow in my passion for mentoring. It has been a privilege and honor to help each of you succeed in your research.

To my first mentor: **Dr. Glenn Legge**, thank you for welcoming me into your lab to conduct undergraduate research. Though I may have been a nagging pest in trying to get the initial interview to work in your lab, thank you for recognizing my perseverance as an asset to your lab. To my little-big-brother **Dr. Roberto Carreño**, thank you for your wonderful friendship and teaching me all things concerning the immune system, SPR, iodoacetamide coupling, silver staining, and protein purification – I miss our wonderful conversations! **Andrea Creath**, where do I begin to thank you for all you have done for me, not just as a researcher, but also for being an amazing prayer partner and for taking me to several doctor's appointments when we all had no idea where my pain was coming from – I love you my dear friend.

To other mentors who have made an indelible impact on my life: Thank you to my high school biology and chemistry teachers, **Mrs. Holbrook** and **Mrs. Boecker** who poured their knowledge into me and ignited my passion for science. To **Dr. Richard Knapp**, you were the first to sit down and have a conversation with me about research, through your encouragement and support letter to Dr. Legge, I have been able to see my dream come to fruition and become the researcher I once only spoke about. **Dr. Christy Landes**, thank you for your infectious ambition - your response to my inquiry as to whether I can and should write a F31 grant, "Why not?! Of course you can!" – was the catalyst to write, apply and submit the grant. I never had the chance to follow up and let you know, "I submitted and I got it!" **Dr. William Lawrence**, thank you for your advice and keeping me apprised of scholarships and grants as an undergraduate

student – without a doubt you spurred my enthusiasm to apply for several funding opportunities thereafter. **Dr. George E. Fox**, thank you for your compassion, empathy and insight into not just research but also life - you have been an amazing mentor, I am going to miss our wonderful conversations. **Dr. Daniel Frigo**, thank you for your invaluable insight and guidance, you have been a terrific mentor in the short time I have known you. **Dr. Ofelia A. Olivero**, your passion for mentoring and research cannot be matched – thank you for your guidance and all you do in lighting the torch for mentoring in science. **Dr. Hye-Jeong Yeo**, thank you for your compassion and friendship – you have been a blessing beyond measure. **Melissa Glueck**, how can I ever thank you for seeing in me what I could not see in myself. Your encouragement and support has been critical to my growth as a mentor and I can't thank you enough for all you have done in providing me with such wonderful opportunities to grow as a researcher and mentor. **Dr. Steven Bark**, you have been such a blessing to me at the darkest time in my life. Thank you for always encouraging me with jokes and making me laugh as you also shared in my silly obsession of awesome 1970-80s movie quotes!

To my future mentor: **Dr. Charis Eng**, the crossing of our paths truly was serendipitous! I am excited to begin this next chapter in my career and for the privilege to learn from you. Thank you for your compassion and grace – I am looking forward to joining my "new family". To my future lab members: **Dr. Sara Akhavanfard**, **Dr. Ritika Jaini**, **Dr. Ying Ni**, **Dr. Blake Chaffee**, **Dr. Mahav Sankunny**, **Janice Blount**, **Todd Romigh**, **Jin-Lian Chen**, **Nick Sarn**, **Jessica Altemus, Weelic Chong**, **Hannah Wang**, **Stetson Thacker**, **Matt Loya**, **Qi Yu**, **and Chad Braley** thank you for making me feel welcomed and receiving me with kindness. **Lamis Yehia**, thank you for your friendship and kind heart – I'm looking forward to great food and partaking in that chocolate fountain!

To my friends and colleagues: **Amanda Paul**, I can't thank you enough for all you did for me at the outset of my matriculation into the Ph.D. program. **Barbara Hey**, a.k.a "Bossy

Barb" what a wonderful gift from God you have been – I thank God every single day for bringing you into my life albeit under the worst circumstances for us both. Thank you for your prayers, shared tears, and blessed hope! I await the day when we will be changed in the twinkling of an eye and will receive our new glorious bodies – hallelujah! To **Blanca Perez**, though we've only known each other for a short time, seeing you on campus and developing a friendship has been a wonderful blessing. **Dr. Ucheena Ossai** – "wow" is all I can say for the amazing person that you are! You have not only been an incredible doctor but an amazing friend – thank you for ALWAYS being there for me when I needed to cry through the pain (or laugh) – you rock! **Mary Wilhite**, you have been a prayer answered – for I prayed months in advance for God to bring a loving, God-fearing family to live next door to us and He surely answered. Thank you for being my go-to prayer warrior and for shining God's Word back to me with encouragement and strength. **Chuck and Debi Hix**, another answer to prayer – Jeff and I prayed for God-fearing neighbors to move in next to us and he brought us the two of you who have been a blessing beyond words! **Rosezelia Jackson**, you have no idea how much your friendship means to me. Thank you for speaking Truth back to me when I was in despair – where would I be without you? **Sabrina Buchanan** and **Percill Griffin** you two have been such a blessing. Thank you for your prayers and support – I will forever cherish our friendship. **Kris Tesh,** thank you for your prayers and for giving me a ride to campus when I couldn't drive – I'm so thankful for our friendship. **Jennifer Myers**, thank you for your kindness and support and for giving me a ride to campus when I couldn't drive. **Josh and Nikki Logsdon**, thank you for being such incredible friends and for all of your support and prayers – though we've only know each other for a short time, I consider y'all as part of my family. **Toni Vychron**, my-sister-from-another-mister, words can't describe how blessed I am to have you in my life. You have been with me in good times and bad – you will always be my sister, I love you! **Yonia Pulido**, thank you for taking the time to help me with the processing of my stipend all these years, but most importantly thank you for your

words of encouragement, they have strengthened me when I needed it most. **Rhonda James**, the crossing of our paths was truly Divine for in you I have found my sister-in-Christ, my prayer warrior, and my rear guard. Thank you for speaking Psalm 40:1-3 over me when I was in need, I will forever thank God for you my dear friend. **Patrick (Abiodun) Bodunrin**, though I've only known you for several months, I know that our friendship was Divinely orchestrated. Thank you for speaking Truth to me and reminding me to "hold fast" 'till He comes. **Dr. Efi Tsoukos**, thank you for your prayers and for being such a wonderful friend – you have inspired me in so many ways. **Dr. Fabiola Mehta**, thank you for being my prayer warrior and sharing encouraging scripture with me when I was bedridden for so long – it was my light in the darkness. **Nanda Karri**, thank you for always making me laugh when I needed it most – you have helped me to see the funny side of life. **Nick Valenzuela**, thank you for encouraging me over the years – never change your zeal for life, it's inspiring to so many. **Drs. Alicia and Lakshmi Bollu**, you two have been such a blessing to me in so many ways. Lakshmi, your passion for science is so inspiring. Alicia, I can't thank you enough for your prayers all these years they've lifted me up beyond measure. **Charlene Henrique**, thank you for taking care of me after my many laparoscopies – and various procedures over the past 20 years. Love you like a sister. To **Bob** and **Sarah Williams**, thank you both for your support, encouragement, and guidance over the years. You both have been such a blessing to Jeff and I and have played a critical role in shaping the individual I am today. To **Anne Geske** and **Karla Anderson**, another Divine encounter – Karla, from the moment we met I knew God's hand was in it – thank you for introducing me to Anne and the both of you bringing food and much needed prayer over to my house after several procedures. May God bless you and your families ever so abundantly. To **Ruben Silva** our neighborhood prayer warrior – thank you for lifting me up before the Heavenly throne over the past year – your prayers moved mountains.

To my physicians: **Drs. Michael J. Heard, Rakesh Mangal, Danielle Antosh, Shaun Lehmann, Ucheena Ossai, Anthony Echo, Pamela Wartian-Smith, Christopher Jayne, Sebastian Faro, Raja Abusharr, Curtis Fandrich, Charles Popeney, and Anna Gonzales**, – each of you have played such a crucial role in my treatment and healing over the years. Without each of you, the completion of this research would not have been possible. Words cannot express how grateful I am. Each of you has gone above and beyond the Hippocratic oath to eliminate my pain – thank you for everything! **Jodi, Sonia, Debbie and Ida**, thank you for praying for me and working me in to see Dr. Heard or Dr. Antosh at a moment's notice.

To my family: **Mom** and **Dad**, I'd like to thank you for bringing me into this world and for encouraging me to climb the tallest mountain – for enriching my life with love and for fostering my love for science and books. Mom, I can't thank you enough for our bi-weekly visits to the library where I always checked out the maximum books that my card would allow. It was in the confines of that small building that I realized my passion for science. Dad, thank you for showing me how to stand strong in the Lord. You both have been my greatest teachers; the values, ethics and principles you taught me are the crux of my foundation. When I look back, I introspect on what really helped me persevere through difficult times. Mom, your demonstration of dedication and resilience, has shaped who I am today – and it is because of you I discovered my love for writing. Dad, through you I understood undeterred perseverance and commitment is required to accomplish anything great. I dedicate this dissertation to you both - it is because of your dedication to our family and your belief in me, that I stand here today, I love you both. **Nira Family**: Thank you to **Marc** and **Cisco Nira**, the best brothers a sister could ever have, you always stood by me, protected me and guided me even when I bossed y'all around (so mom says – lol!), it is because of y'all's encouragement and support, I am who I am today – I love you both more than you'll ever know. To **'Cole Nira**, my beloved sister, how can I say in such a small space what you mean to me, when you came into this world, I knew that my life would forever be

changed, because from that moment on, I was responsible for something greater than myself. I raised you like you were my own daughter, took you everywhere with me – you helped shape my life. Thank you for always standing by my side, for your prayers, strength and encouragement – I love you. To **Michelle Nira**, **Nikki Nira**, and **Angela Ruth** having y'all as sister-in-laws has been such a blessing to me in more ways than y'all could ever imagine. Thank you for your prayers, your love, and encouragement. To **Aunt Carmen**, **Aunt Tootsie,** and **Rita Nira**, thank you is not merely enough for how I wish to express my gratitude towards y'all - not just your prayers of encouragement but also for driving from afar to take care of me (and Jeff) during my most difficult surgery – I love you all so very much. **Smith Family**: To **Sandy** (mom), **Joe** (dad) and **Lisa Smith**, thank you for treating me like a daughter, for your unconditional love and for all of your prayers and support that has sustained me all these years – I love you all so much. To my church family**: Angel Smith, Robin Beale, Dr. Kinglsey** and **Chinyere Nzeadibe, Dr. Shaun** and **Chelsea Lehmann, Pastor Chris** and **Franchesca Jones, Ken** and **Arlene Nelson, Jane Ramos** and **Chris Nessing, Etalia Castillo, De Demouy, Terrance and Angel White, Andrew and Lily Hernandez, Sue Sigarroa, Kieron** and **Sasha Prince, Annya de Gracia, Merlyn Maurice, Victoria Simons, Debra Washington, Ben** and **Pat Cartwright, Paula Thomas, Cara Fandrich, Dennis** and **Carolyn Lehmann, Andrew Gencer, Bonnie** and **Mike Conlin, Sandy** and **Robert Monk, David Pershall, Carolina Cirilo,** and so many others at The Adventist Church of the Woodlands and Central Seventh-day Adventist Church – y'all have carried me in so many ways, and I am forever grateful. It is because of your steadfast prayers and love that I made it through the fire. Thank you for caring for me (and our dogs) when I was in such dire pain, thank you not only for your encouragement to me but also to Jeff. So many of you have gone before His throne seeking healing, comfort and peace on my behalf – praise God for His faithfulness, for He has gone above and beyond what you could ever think or ask. Thank you for the preparation and delivery of meals and beautiful floral arrangements during my many

surgeries. A special thank you to **Dr. Shaun** and **Chelsea Lehman** for taking me to the ER multiple times, for the preparation and delivery of food as well as the administration of much needed and beneficial medication. **Chelsea**, thank you for your encouragement in my most despairing times. And thank you for taking the time to review and edit this dissertation – you are such a blessing!

To my husband **Jeff Smith**, I could barely begin typing this section before I started crying…again – It is because of you, that I can write this and I dedicate this dissertation to you. I can't think of words strong enough to express my deep love and gratitude towards you. You have seen me at my absolute lowest all the while never leaving my side. The countless doctors visits, surgeries, ER visits, medication, new and often frightening diagnoses, the weeks, sometimes months that I was bedridden – you have lived the definition of "in sickness and in health" and continue to amaze me with your support and encouragement. There is nothing I could ever do to repay you for all that you have done and continue to do for our family. Thank you for helping me to keep my focus on God and for making me laugh when I trembled in fear every time I was poked, prodded, injected with medication or being wheeled into the operating room. Thank you for your patience when I was in pain and not a "good patient". Thank you for revealing and sharing your heart with me – for standing by my side all these years while I agonized in pain – expecting that it would go away and holding on to the hopes and dreams of ever bearing a child only to realize that endometriosis has forever robbed me of that blessing. Thank you for being my rock when I had no strength, and for praying over me when I was discouraged. I love you and pray that God continues to strengthen our marriage and blesses us in this exciting and new chapter in our lives.

To my babies: **Kya and 'Koda** - you two have been at my side through it all. Y'all have literally wiped my tears away and filled my heart with joy when all I could do was cry. I praise God for your unconditional love. Y'all have given me hope when I felt hopeless, happiness when

I felt despair and revealed so much to me about life, love and loyalty. Y'all have taught me that being a mother doesn't mean being related to someone by blood – it means loving unconditionally and with all your heart.

I gratefully acknowledge the funding that was provided for the entire duration of this research from the National Institute of Health and National Cancer Institute (NIH/NCI). Supercomputer time was provided by the Center for Advanced Computing and Data Systems (CACDS) as well as Extreme Science and Engineering Discovery Environment (XSEDE).

And I haven't forgotten, to ***all the women*** who have been or will be diagnosed with endometriosis, I stand with each of you in the uncertainty of your pain – my hope is that this research will lay the foundation for improved diagnostic techniques and treatment of endometriosis. This disease will not win; it will no longer steal the hopes and dreams of our lives!

UNDERSTANDING STRUCTURE AND DYNAMICS OF PTEN AND ITS POSSIBLE GENOTYPE-PHENOTYPE CORRELATIONS IN ENDOMETRIOSIS AND CANCER

An Abstract of a Dissertation

Presented to

the Faculty of the Department of Biology and Biochemistry

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Iris Nira Smith

December 2016

Abstract

The phosphatase and tensin homolog deleted on chromosome 10, (*PTEN)* gene encodes a tumor suppressor phosphatase frequently mutated in various human cancers. Somatic missense mutations of PTEN have recently been found in patients with endometriosis, endometrial cancer, and ovarian cancer. Here we present the first computational analysis of 13 somatic missense PTEN mutations to assess a possible genotype-phenotype correlation in endometriosis and cancer. We posit PTEN's active site defines a possible mutation-driven allosteric region wherein a subset of mutations correlate with endometriosis, endometrial cancer, and ovarian cancer. Our data suggest that mutations within the active site disrupt the structural stability, electrostatic interaction, global dynamics and the structural communication pathway, likely contributing to the aforementioned phenotypes.

Multiple *in silico* prediction methods were utilized to calculate protein structural stability changes produced by each mutation; decreases in protein structure stability were seen in each mutation with an increase in dynamics across the phosphatase-C2 domain interface of R130G/L/Q and R173C/H mutations. To assess the impact on intrinsic and global dynamics, elastic network models (ENMs) were employed demonstrating changes from wild-type "hinge-bending" to "zipper-like" global motions induced by each mutation. All-atom molecular dynamics (MD) simulations revealed large conformational changes that affect the global dynamics of the active site loops and the CBR3 loop in the C2 domain. Interestingly, mutations G36E/R, C124S, G129R, R130L/Q, R173C/H, and V191A dramatically affected the principal motions of the active site loops and inter-domain interface. Overall, the global dynamics induced by each mutation effects reveal unique long-range perturbations that may impair PTEN's function.

We further investigated structural communication within each mutant system using protein structure network (PSN) analysis and found that R130 and R173 play critical roles in controlling salient communication pathways suggesting a compelling interplay between the two positions involving a potential mutation-driven allosteric interface. The results of this research provide a greater understanding of the mechanistic role of mutated PTEN associated with endometriosis and cancer. It is our hope that these results will aid in a better clinical-molecular classification of the resulting phenotypes allowing for translation into improved diagnostic and therapeutic approaches.

Table of Contents

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Abbreviations

ANM	Anisotropic network model
APBS	Adaptive Poisson-Boltzmann solver
ENCoM	Elastic network contact model
ENM	Elastic network model
ESP	Electrostatic surface potential
GNM	Gaussian network model
GROMACS	Groningen machine for chemical simulations
LINCS	Linear constraint solver
MD	Molecular dynamics
NMA	Normal mode analysis
PCA	Principal component analysis
PDB	Protein data bank
PIP_3	Phosphatidylinositol-3,4,5-triphosphate
PME	Particle mesh Ewald
PSN	Protein structure network
PTEN	Phosphatase and tensin homolog deleted on chromosome ten
Rg	Radius of gyration
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
SASA	Solvent accessible surface area
UHBD	University of Houston Brownian dynamics
VMD	Visual molecular dynamics
WT	Wild-type

Chapter 1: Introduction

1.1 Phosphatase and tensin homolog deleted on chromosome ten, (*PTEN*)

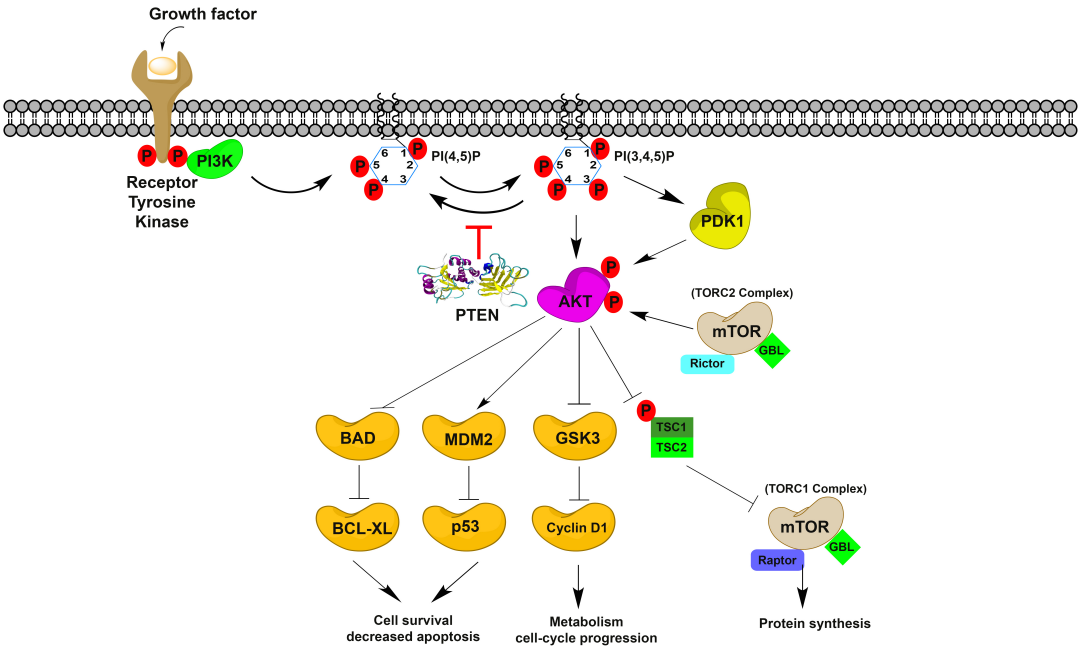
1.1.1 PTEN-PI3K signaling

Originally discovered in 1997, the *PTEN* (phosphatase and tensin homolog deleted on chromosome 10) tumor suppressor also referred to as MMAC (mutated in multiple advanced cancers) and TEP-1 (TGF-β-regulated and epithelial cell enriched phosphatase) was the first phosphatase identified to be frequently mutated/deleted in various human cancers.[1-3] The *PTEN* gene, located at 10q23.3, encodes a 403-amino acid dual-specificity phosphatase having both protein and lipid phosphatase activity.[4] As a protein phosphatase, PTEN exerts its function by dephosphorylating tyrosine-, threonine-, and serine-, phosphorylated proteins thus regulating the cell cycle and proliferation.[4] However, its most relevant function is based in its ability to catalyze the dephosphorylation of the 3' phosphate of the inositol ring in phosphatidylinositol (3,4,5)-triphosphate (PIP_3), an important intracellular lipid second messenger, with greater alacrity, thus lowering it's level within the cell.[5]

In dephosphorylating PIP_3 to produce phosphatidylinositol (4,5)-bisphosphate (PIP_2), PTEN negatively regulates the upstream phosphotidylinositol 3-kinase (PI3K)/Akt pathway thus restraining downstream anti-apoptotic and growth stimulatory effects of PDK1 (phosphoinositide-dependent kinase) and protein kinase B (PKB/Akt) that promote cellular proliferation and survival.[4,6] The PI3K/Akt pathway is a major survival pathway activated in cancer with PTEN constituting the main node of inhibition by limiting the phosphorylation and activation of Akt, thereby regulating progression of the cell cycle and induction of apoptosis.[7]

Hence, the molecular mechanisms attributed to PTEN are consistent with its role as a tumor suppressor.

Figure 1.1 PTEN-PI3K-AKT signaling pathway



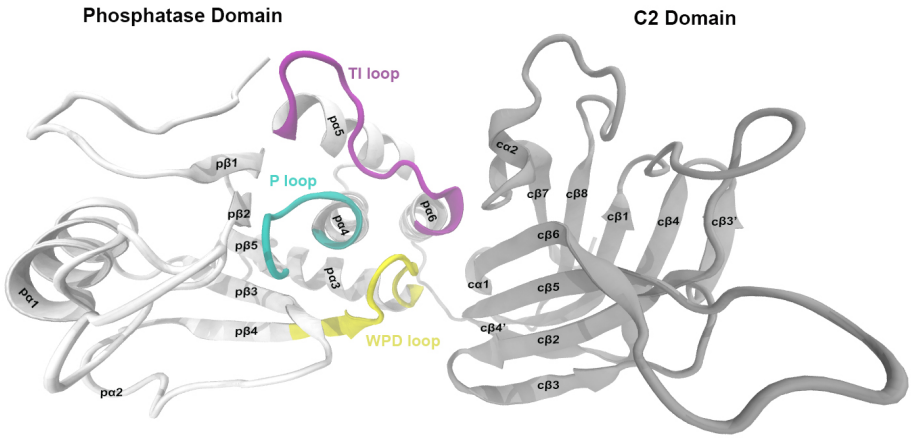
1.1.2 PTEN structure

The human PTEN protein consists of 403 amino acid residues with multiple domains that contribute to its function and stability which include: (i) phosphatidylinositol 4,5-bisphosphate – binding motif (residues 6-15), (ii) N-terminal phosphatase domain (residues 15-185), (iii) C2 domain (residues 185-351), (iv) C-terminal tail (residues 352-400), (v) two PEST motifs (residues 350-375 and 379-386), (vi) PDZ-binding domain (residues 401-403), (vii) and two ATP-binding motifs (residues 60-73 and 122-136).[4,8-12]

Figure 1.2 PTEN domain structure

The full length PTEN protein, however was not amenable to crystallization, and proteolytic digestion of N-terminal residues 1-13, internal D-loop (residues 286-309), and C-tail (residues 353-403) revealed these regions as loosely folded or unstructured, thus, identifying PTEN as an intrinsically disordered protein (IDP).[4,13] The functional versatility of PTEN is attributed to its IDP regions which allow it to interact with over 400 proteins in different subcellular compartments such as plasma membrane, cytoplasm, nucleus, and exosomes.[14] However, recombinant PTEN lacking unstructured regions resulted in the crystallization of PTEN (PDB ID 1D5R) which include only the phosphatase domain (residues 15-185), that accommodates the invariant signature motif HCXXGXXR (commonly abbreviated $C(x)_5R$) and a tightly associated C2 domain (residues 186-351) which participates in membrane binding.[4]

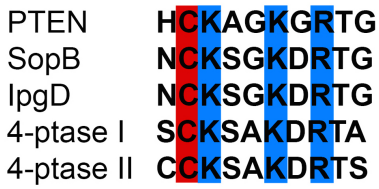
Figure 1.3 PTEN three-dimensional structure



The phosphatase domain contains a central five-stranded β-sheet with six α-helices (two α-helices on one side and four on the other). Similar to other protein tyrosine phosphatases (PTPs) and dual-specificity phosphatases (DSPs), PTEN has two α-helix motifs that flank the catalytic active site that consist of: (i) α1-helix (residues 48-59) and (ii) α6-helix (residues 161-177).[15,16] Though not directly involved in catalysis, the α1-helix assists with the formation of the overall secondary structure.[15,16] The C2 domain represents a β–sandwich that consists of two antiparallel β-sheets with two short α–helices intervening between the strands and three CBR3, Cα2, and Cβ1/2 loops located on the membrane facing surface. The C2 domain bears hallmarks of Ca^{2+}-dependent phosphatases, however, it lacks the essential Asp-268, a Ca^{2+} ligand and therefore binds to membranes in a Ca^{2+} -independent manner.[4]

The phosphatase and C2 domains associate across an extensive interface that is adjacent to the active-site and consists of conserved residues frequently mutated in cancer.[4] The interactions of these two domains are controlled by three loops located in the catalytic active-site pocket. The wall of this pocket is delimited in part by the WPD loop (residues 88-98), the signature motif P loop (H^{123}CKAGKGR^{130}), and TI loop (residues 160-171) which contain residues that are responsible for catalysis (D92, C124, and R130), the overall positive charge within the active-site (H93, K125, K128), mediation of loop motion (H123 and G127), and govern the depth and width of pocket (4-residue insertion K163, K164, G165, V166).[4] The phosphatase domain shares homology with both tensin and auxilin proteins and contains two signature motifs: (i) the phosphate binding loop (P loop) HCXXGXXR and (ii) the "Trp-Pro-Asp" loop (WPD loop) commonly found in the regulatory sites of PTPs and DSPs (**Figure 1.4**).

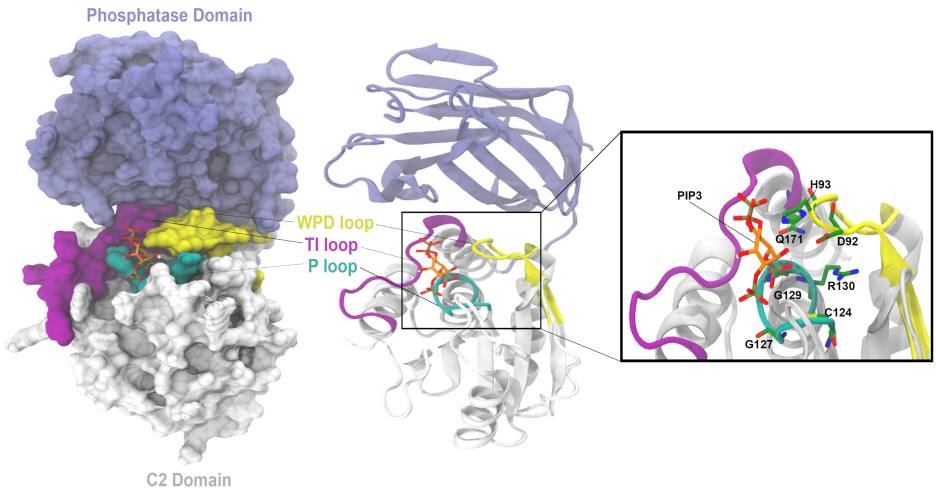
Figure 1.4 PTEN signature motif alignment



Adapted from Maehama and Dixon 1999[17]

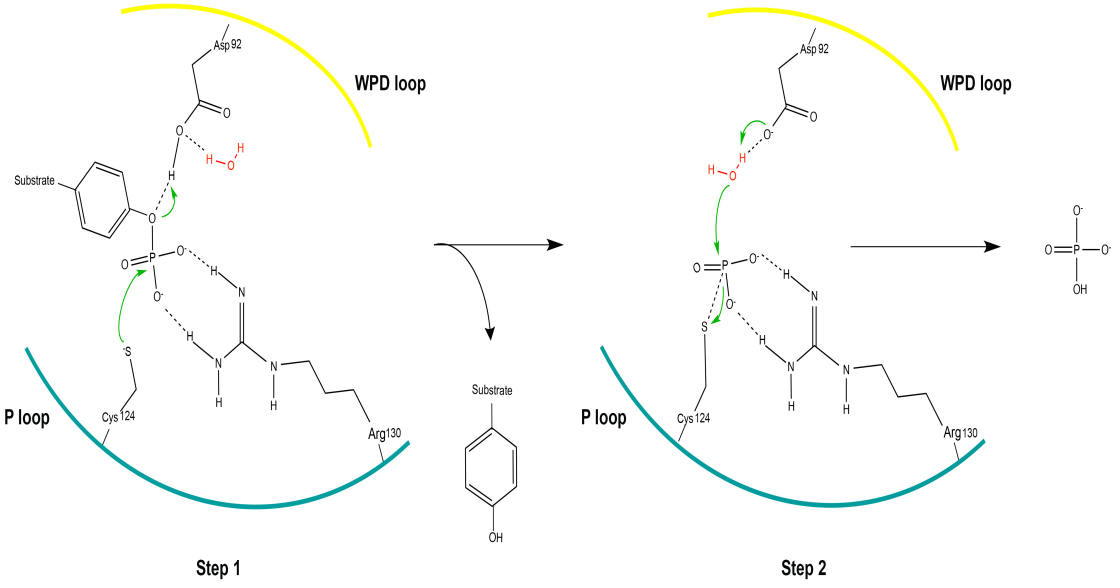
The phosphatase domain however contains a rather wide and deep catalytic pocket specific only to PTEN. In fact, the architecture of the active-site of PTEN is sufficiently large enough to accommodate the sugar head group of inositol phospholipids as a substrate allowing for catalytic specificity.[4,17-19] The active-site pocket of PTEN is ~8 Å deep with an elliptical opening of ~5 x 11 Å.[4] The extension and larger width of the active-site pocket is due to the four-residue insertion of the TI loop (**Figure 1.5**). The P loop contains residues K125 and K128 at its center, which along with H93 of the WPD loop imparts a highly positive charge to the pocket leading to the preference of negatively charged PI(3,4,5)P_3 and highly acidic polypeptide substrates.[4] The pocket extension and highly positive charge of said pocket coupled together set PTEN apart from other phosphatases. Mutations in the TI loop (T167), WPD loop (H93 and D92), and P loop (C124, K128, G129, and R130) lead to a reduction in PIP_3 phosphatase activity by ~75%.[4]

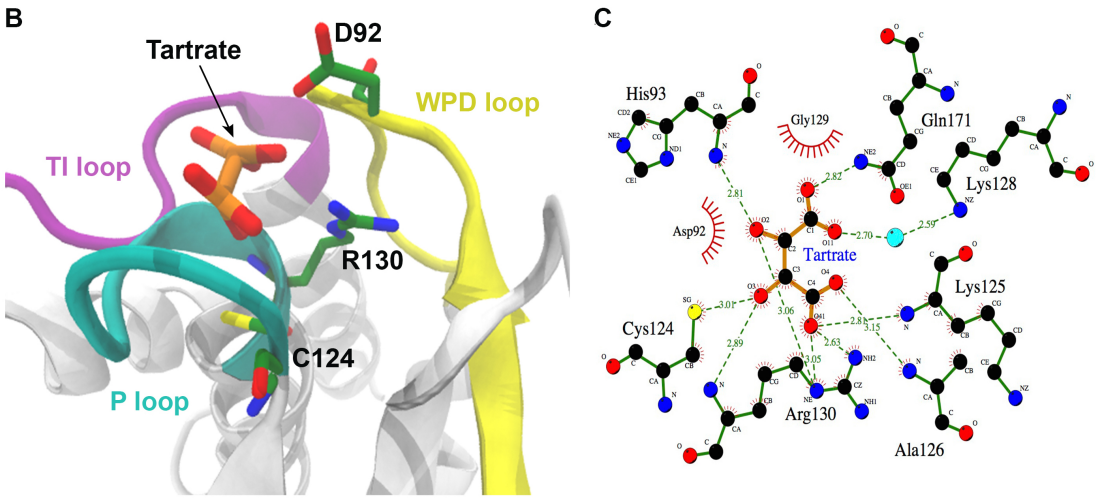
Figure 1.5 PTEN-PIP$_3$ structure



1.1.3 Mechanism of action

Figure 1.6 PTEN catalytic mechanism





Adapted from Zhang[20] and Tautz[21]

As a dual-specificity phosphatase, PTEN can dephosphorylate both Ser/Thr/Tyr protein substrates as well as phospholipids substrates. The mechanism determining whether PTEN acts as a protein or lipid phosphatase remains unclear. However, PTEN is highly selective for the lipid second messenger, PIP_3 and catalyzes its dephosphorylation via a two-step mechanism: (i) through the transfer of the target phosphate moiety from the substrate to the conserved cysteine forming a cysteinyl phosphoenzyme intermediate, and (ii) hydrolysis of the phosphoenzyme intermediate to form inorganic phosphate.

Specifically, the P loop contains both the catalytically conserved cysteine at position 124, which forms a cysteinyl phosphoenzyme intermediate that is subsequently hydrolyzed, and arginine 130, which binds to the D3-phosphate of PIP_3 transferring it to C124. In step 1, the catalytic cysteine, located at the base of the active-site cleft, acts as a nucleophile by attacking the phosphorous atom in the phosphate moiety of PIP_3 resulting in the formation of a thiol phosphate intermediate.[22,23] Arginine 130 plays a crucial role in the binding of the phosphoryl group contributing to transition-state stabilization, [23-25] while D92, a general acid of the WPD loop, donates a proton to the substrate-leaving group. In step 2, D92 now serves as a general base by

abstracting a proton from a water molecule. The cysteinyl phosphoenzyme intermediate is then hydrolyzed resulting in the liberation of the PO_3 moiety and enzyme regeneration.[20,25,26] The conserved residues histidine 123 and threonine 131, assist in regulating both the pKa of C124 as well as in facilitating hydrolysis.[25]

The conserved aspartic acid 92 (~30-40 residues away from catalytic cysteine) in the WPD loop acts both as a general acid and general base, respectfully, in step 1 and step 2 of the dephosphorylation mechanism.[27] Moreover, the WPD loop may undergo rapid movement upon substrate binding allowing for optimal positioning of D92 for catalysis.[27,28] The WPD loop is designated in the "open" state prior to substrate binding to the P loop signature motif where it is oriented away from the active-site.[29] Upon substrate binding, the WPD loop moves toward the catalytic pocket, bringing D92 in close proximity to the substrate.[29] This conformation is designated as the WPD loop "closed state". Dephosphorylation of phospholipid substrates and phosphopeptide substrates reveal a pronounced difference in the role of D92 in the dephosphorylation of the two types of substrates.[27] In fact, phospholipid substrates and not phosphopeptide substrates are able to recruit the WPD loop into the active-site such that D92 can participate in catalysis. Previous studies of the recombinant PTEN in complex with tartrate, which mimics the phosphorylated inositol ring of PIP_3 (PDB ID 1D5R), illustrates the WPD loop adopts the "closed" conformation.[27] It is posited that the WPD loop is recruited to the active-site by way of H93 via interaction with the D5 phosphate of PIP_3.[4]

1.2 Endometriosis

1.2.1 Overview and definition

Endometriosis is a chronic estrogen-dependent inflammatory disease defined as the presence of endometrial tissue outside (ectopic) the uterine cavity in various areas throughout the

body developing into lesions, adhesions, implants, or tumors resulting in severe dysmenorrhea (painful menses), dyspareunia (painful intercourse), dysuria (painful urination), dyschezia (defecation difficulties), chronic fatigue, frequent miscarriages, and infertility.[30-34] Ectopic endometrial tissue (uterine womb lining) located within the muscle wall of the uterus is called endometriosis interna or adenomyosis. The extent of endometriosis varies from small red, blue/black, or white lesions on the outer surface of the pelvis, peritoneum or pelvic organs to large ovarian endometriotic cysts (endometriomas) that invade the ovaries.[35] This displaced endometrial tissue responds to hormonal changes during menses resulting in irritation and internal bleeding which produces extensive fibrosis and internal scar tissue known as adhesions that spread out across pelvic organs, binding them together, leading to distortion of the pelvic anatomy. Occasionally, endometriosis can be found on the kidneys, bladder, diaphragm, liver, nerves, ureter, bowel, walls of intestine, in tissue between the vagina, and rectum, and even the brain.[35-46]

1.2.2 Statistics

Endometriosis affects an estimated 176 million women worldwide showing no disparity towards age, ethnicity, or social circumstances.[47] With an estimated frequency of 5%-10% among women of reproductive age[31,33,34], endometriosis is the third leading cause of gynecologic hospitalization in developed countries and a leading cause of hysterectomy.[31] The prevalence rate of pelvic endometriosis approaches 6-10% in the general female population; in women with pain, infertility, or both, the frequency is 35%-50%.[33]

1.2.3 Pathogenesis

Though the exact pathogenesis and etiology of endometriosis has yet to be elucidated,

there are several mechanistic theories as to how the endometrial tissue escapes the uterine cavity and implants itself onto organs of the pelvic cavity (**Table 1.1**). Leading theories include: (1) ***Sampson's theory*** – retrograde menstrual reflux through the fallopian tubes into peritoneal cavity,[48] (2) ***Metaplastic theory*** – conversion of coelomic epithelium into various types of tissue,[48] (3) ***Embryonic rest müllerianosis theory*** – remnant müllerian cells that remained in pelvic tissue during development of the Müllerian system. Under estrogen stimulation, residual cells from embryologic müllerian duct migration may be induced into functioning endometrial glands and stroma.[48] (4) ***Benign metastasis theory*** – ectopic endometrial implants are the result of lymphatic or hematogenous dissemination of endometrial cells into the venous circulation.[49,50] However, more recently, it has been accepted that endometriosis has a multifactorial etiology which include ***genetic, hormonal,*** and ***environmental factors***.[33,48,51]

Sampson's theory is the oldest and most accepted theory explaining the etiology of endometriosis. This theory suggests that endometriosis occurs as a result of the retrograde flow of sloughed endometrial cells/debris via the fallopian tubes into the pelvic cavity during menstruation.[52] However, retrograde menstruation through the fallopian tubes occurs in 70-90% of women and not all of these women develop endometriosis.[53] Yet, patients that develop endometriosis have a larger volume of retrograde menstrual fluid found in their pelvises compared to healthy women, which may increase the risk of endometriotic lesion implantation.[54]

Metaplastic theory may explain the occurrence of endometriosis in pre-pubertal girls.[55] Metaplasia of the coelomic epithelium involves the transformation of normal peritoneal tissue to ectopic endometrial tissue induced possibly by environmental factors that promote the differentiation of cells resulting in the formation of endometrial cells and endometriomas.[33,56] However, the driving force for endometrial growth, estrogen, is absent in pre-pubertal girls, and therefore the genesis of endometriosis may be different in women of reproductive age.[55] The fact that ectopic endometrial tissue has been detected in female fetuses as well as in men, suggests

that endometriosis may be the result of defective embryogenesis.[57,58]

The ***Embryonic rest müllerianosis theory*** best explains the implication of endometriotic lesion development under the influence of estrogen or estrogen mimetics. According to this theory, residual Müllerian or Wolffian ducts grow caudally, fuse with the urogenital sinus of the pelvis, and develop into endometriotic lesions that respond to estrogen.[56] ***Benign metastasis*** postulates that endometrial cells gain entry into open basal lymph and blood vessels and are embolized to ectopic sites. The growth of endometriotic lesions could theoretically seed additional lesions in the pelvic cavity and other distal sites.[59]

More than 20 years ago heritable ***genetic factors*** were recognized when the risk for first-degree relatives of women with severe endometriosis were reported to be six times higher than that for relatives of unaffected women.[60-63] In fact, familial aggregation has been shown in twin studies[64-66] and in clinical[60,67] and population-based[68] samples. Additionally, various groups have reported candidate genes that have potential biological susceptibility to endometriosis by the utilization of linkage analysis and affected sibling pairs (**Table 1.2**),[69-76] this includes genes associated with malignant transformation (PTEN, p53, estrogen-, progesterone-, and androgen-receptors).[76-79]

Since endometriosis is an estrogen-driven disease, the role of steroid ***hormones*** plays a central role in the pathogenesis of endometriosis. Similar to the eutopic (normal) endometrium, endometrial proliferation, and the growth of ectopic lesions have an increased response to estrogen thus enhancing the progression of endometriosis.[80] Moreover, endometriotic tissue relative to eutopic endometrium exhibits an increased expression of the aromatase enzyme and decreased expression of 17β-hydroxysteroid dehydrogenase (17β-HSD type 2).[81] Aromatase gives rise to local biosynthesis of estradiol thereby stimulating production of prostaglandin E_2 and establishing a positive feedback cycle.[81,82] Prostaglandin E_2 further stimulates activity of aromatase favoring the accumulation of estrogen and prostaglandins. Thus, estradiol synthesis is

increased while its inactivation is decreased, resulting in higher concentrations of the hormone and increased proliferation of endometriosis.

The increasing incidence of endometriosis as well as higher prevalence rates of severe endometriosis in industrialized countries suggests a possible link between endometriosis and ***environmental factors***. However, to date, there has been no large-scale epidemiological study definitively linking a particular class of chemicals to the risk of endometriosis.[33] Although estrogen-like compounds,[83] endocrine disruptors,[84] and the environmental toxin dioxin[85] have been suggested as possible factors, the developmental timing of action of such agents and their roles in influencing other systems that predispose endometriosis must be considered in the context of genetic background as well as stimulus-driven reprogramming of the female reproductive tract.[56]

Table 1.1 Theories on the pathogenesis of endometriosis

Theory	Explanation
Mechanistic Theories	
(1) *Sampson's theory* – retrograde implantation	Endometrial tissue sloughed through fallopian tubes into peritoneal cavity
(2) *Metaplastic theory* – coelomic metaplasia	Extrauterine cells that abnormally differentiate into endometrial cells
(3) *Embryonic rest müllerianosis theory*	Remnant müllerian cells that may be induced into endometrial glands and stroma under estrogen stimulation
(4) *Benign Metastasis theory*	Lymphatic and vascular dissemination of endometrial cells
Multifactorial Etiologies	
(5) *Genetic basis*	Alteration of cellular function that increases the attachment of ectopic endometrial cells to the peritoneal epithelium
(6) *Hormonal basis*	Estrogen-driven proliferation of endometrial lesions
(7) *Environmental basis*	Exposure to toxic chemicals (i.e. dioxins, PCBs) and endocrine disrupting chemicals

Table 1.2 Candidate genes and susceptibility to endometriosis

Gene	Chromosomal locus
Steroid Enzymes and Hormone Receptors	
Cytochrome P450 1A1 (CYP1A1)	15q24
Progesterone Receptor (PGR)	11q22-23
Androgen Receptor (AR)	Xq12
Hydroxysteroid (17-β) dehydrogenase 1 (HSD17β1)	17q11-21
PPARγ2 Pro-12-Ala allele	3q25
Estrogen Receptor	
ERα (ER1)	6q24-27
ERβ (ER2)	14q21-22
Xenobiotic Metabolism	
N-acetyl Transferase 2	8p22
Glutathione-M-Transferase 1 (GSTM1)	1p13.3
Glutathione-S-Transferase 1 (GSTT1)	22q11.2
Signal Transduction	
Phosphatase and Tensin Homolog deleted on chromosome ten (PTEN)	10q23
Kirsten Rat Sarcoma viral oncogene homolog (KRAS)	12p12
Tumor Suppressor - TP53 (p53)	17p13
Adhesion and Inflammation Molecules	
Interleukin adhesion molecule (ICAM1)	19p13
Tumor Necrosis Factor-α (TNFα)	6p21.3
Vascular Endothelial Growth Factor (VEGFA)	6p21-12
Glutathione-1-phosphate uridyl Transferase (GALT)	9p13
Transcription Factor	
Empty Spiracles Homeobox 2 (EMX2)	10q26.1

1.2.4 Diagnosis

Establishing a diagnosis of endometriosis can be quite challenging with the time between onset of symptoms and a definitive diagnosis in most cases taking several years. In fact, the average delay of diagnosis is approximately 7-12 years.[86-91] Non-invasive techniques have been utilized to detect endometriosis namely, magnetic resonance imaging (MRI),[92] computed

tomography scans,[93] ultrasound,[94-99] and transvaginal hydrolaparoscopy[100] as well as the use of serum and peritoneal fluid biomarkers such as cancer antigen 125 (CA-125),[101,102] cytokines Interleukin-6 (IL-6),[103] and tumor necrosis factor-alpha (TNF-α).[103] However, further investigation is needed to evaluate the relevance and validity of these non-invasive techniques for diagnosis, as these modalities are not helpful in assessing the severity and invasiveness of endometriosis on a case by case basis.

A definitive diagnosis can only be made by means of surgical laparoscopy which is considered the gold standard and primary diagnostic modality for endometriosis. Furthermore, the severity of endometriosis is determined by the American Society of Reproductive Medicine (ASRM) classification system which quantifies disease by a point system into one of four stages (**Table 1.3**): minimal disease (**stage I**), mild disease (**stage II**), moderate disease (**stage III**), and severe disease (**stage IV**).[104] Through laparoscopic visual observations, disease severity is classified based on location, depth of invasion, extent of endometriosis implants, severity of adhesions, and presence/size of ovarian endometriomas.

Table 1.3 Stages of endometriosis

Stage	Level of Severity	Description	Total Points
I	Minimal	Presentation of 2-3 brownish, reddish, blue-black, white or clear superficial implants	4
II	Mild	Appearance of more implants that occur within deeper layers of tissue	9
III	Moderate	Many deep implants in combination with minor/small endometriomas on one or both ovaries. May have chocolate cysts and filmy adhesions	26-30
IV	Severe	Persistence of deep implants, enlargement of endometriomas on one or both ovaries, development of dense adhesions. Presence of chocolate cysts. Reproductive organs are bound down by growths, bladder, and/or bowel may also be affected	52-114

1.2.5 Treatment

Although there is currently no known cure, hormonal and surgical options are available to manage symptoms associated with the disease. Endometriosis progresses and represses in an estrogen-dependent fashion; therefore hormonal-medical treatment provides the basis for therapeutic interventions. **Hormonal treatment** with a combination of oral contraceptive pills (OCPs), progestins, gonadotropin-releasing hormones (GnRH) analogues, danazol, and the levonorgestrel-releasing intrauterine system (lng-IUS) is warranted to repress growth and proliferation of endometriotic lesions and alleviate pain. Yet treatment with lng-IUS, only alleviates pain and does not affect repression of disease.[105] Aromatase inhibitors either alone or in combination with oral contraceptives are also beneficial in reducing pain symptoms.[81,106-108] Non-steroidal anti-inflammatory drugs (NSAIDs) are also widely utilized to treat chronic pain associated with endometriosis. However, sometimes more powerful medications like opioids are needed. Additional pain options include acupuncture, physiotherapy, and the use of transcutaneous electrical nerve stimulations (TENS).

Surgical treatment for endometriosis involves the excision of endometriotic lesions, cysts, and fibrotic adhesions lowering the inflammatory response, reducing pain and improving fertility. Surgical treatment is broadly classified as conservative (excision of cysts, endometriotic lesions, and adhesions), semiconservative (removal of uterus, cervix, and one or more ovaries) and radical (removal of uterus, cervix, fallopian tubes, ovaries, and vagina).[109] In addition to excision surgery, a pre-sacral neurectomy (PSN), endometrial ablation, or laparoscopic uterine nerve ablation (LUNA) can be performed to aid in pain management.[110,111] Definitive surgery which includes hysterectomy and removal of ovaries is reserved for women with intractable pain or who no longer desire pregnancy.[112] However, a high recurrence rate of 62% has been reported in advanced stages of endometriosis, in which ovaries were conserved after hysterectomy.[112,113]

1.2.6 Current challenges in diagnosis and effectiveness of treatment

Despite decades of research and significant understanding of endometriosis, the molecular mechanism and pathogenesis underlying its proliferation remains incompletely understood. The lack of a reliable non-invasive diagnostic test contributes considerably to an extended delay between the onset of symptoms and arriving at an accurate diagnosis of endometriosis (~7-12 years). The diagnostic challenge is further compounded by the unreliable correlation between clinical manifestations and surgical findings[114] – as the ASRM clinical staging system assesses physical disease only and does not correlate with the severity of pain that patients experience. Often times patients with mild to minimal disease (stages I-II) experience excruciating pain, while those with severe disease (stages III-IV) may have no symptoms at all.

However, establishing diagnoses on the basis of symptoms alone, can be complicated as women with endometriosis also have a high incidence of co-existing conditions, including allergies, asthma, fibromyalgia, irritable bowel syndrome, pelvic floor dysfunction, autoimmune disease, hypothyroidism, interstitial cystitis, chronic fatigue syndrome, and a low resistance to fevers.[91] Moreover, as symptoms increase in severity, quality of life is further reduced owing to the significant negative impact on social, familial, sexual, educational, and professional aspects of daily life.[88,115-117]

From a societal perspective, the economic burden from endometriosis is estimated at $119 billion/year (U.S.) for both direct and indirect costs, comparable to other chronic diseases though endometriosis is far less understood and often socially unaccepted.[116,118] This substantial price tag assumes a 10 percent prevalence rate among women of reproductive age in the U.S., not reflecting the potential global financial burden with a total of 176 million women affected worldwide. Loss of productivity at work/school (~10-12 hours/week), involvement of multiple healthcare professionals, diagnostic and surgical procedures, “hit and miss” treatments, as well as costs of medicines, all factor in the direct and indirect costs of endometriosis to patients and

society. As a result, the average annual total societal cost per woman is estimated to be $16,116.[118] In the U.S. alone, the associated costs of endometriosis has grown substantially from $15-22 billion in 2002, with a significant portion of that cost attributable to the expensive procedure of diagnostic laparoscopy.[116,119-122]

Although hormonal suppression is an effective first-line approach for treating pain, it is not without risks or adverse side effects, as pharmacological agents are often of limited efficacy and counterproductive to fertility. Moreover, hormonal treatments have no long-term effect on endometriosis, therefore, surgical removal of endometriotic lesions is recommended and aims not only at reducing pain but is also considered the standard in treating endometriosis. Regardless of treatment, endometriotic lesions recur in more than 30% of patients, estimated at 40-50% within 5 years, oftentimes with reduced fertility and increased pelvic pain.[123,124] Despite the difficulties in distinguishing between recurrence and persistence of endometriosis,[48] it is of utmost concern to monitor these patients over the years as their risk of cancer increases with long-standing disease.[125-129]

1.3 Linking PTEN to endometriosis and cancer

1.3.1 Endometriosis and malignant transformation

Endometriosis is considered to be a benign condition, however, it bears hallmarks of cancer such as development of local and distant foci and attachment to and invasion of other tissues with subsequent damage to targeted organs. The malignant transformation of endometriosis was reported as early as 1925,[52,130,131] with a number of studies later identifying carcinomas arising from endometriosis of the colon,[132,133] the bladder,[134,135] the vagina,[136,137] the vulva,[138] and the abdominal wall.[139-141] Research suggests that women with endometriosis may experience elevated risks of a variety of different types of cancer.[126,127,129,142]

Though early studies suggest a link between endometriosis and cancer,[52,126,131,143,144] an exact frequency of malignant transformation is difficult to ascertain, however various values are reported in the literature.[130] Studies have long suggested that ovarian and endometrial carcinoma arise through malignant transformation in 15-40% and 1% of all endometriosis cases, respectively.[50,127,130,144-156] In fact, women with a prolonged history of ovarian endometriosis (>10 years) and a diagnosis before age 30, had a higher relative risk for ovarian cancer.[126]

1.3.2 Endometriosis and endometrial cancer risk

Although association between endometriosis and some subtypes of ovarian cancer has been well established, the association between endometriosis and endometrial cancer is not as thoroughly defined. However, an increased association between endometriosis and endometrial cancer has been established and is typically found in woman with advanced stage endometriosis (stage IV).[157] Histological and immunohistochemical analyses classify endometrial cancer into two broad categories: **type I** (endometrioid (EMC) – variants: secretory EMC, endocervical adenocarcinoma, mixed-type adenocarcinoma) and **type II** (serous carcinoma, clear-cell carcinoma, and carcinosarcoma).[158] The most common genetic alterations in type I include microsatellite instability and mutations in *K-RAS*, *β-catenin*, *PTEN*, and *PIK3CA* genes.[159] In contrast, type II endometrial serous carcinoma, demonstrates mutations in *TP53* gene and HER2 overexpression or amplification.[160,161] Type I accounts for ~80-85% of cases, whereas type II encompasses 10% of all endometrial cancer cases.

Clinically, endometriosis has been identified in ~30% of cases with synchronous endometrioid type endometrial and ovarian cancers.[155,162-164] In fact, 50% of women with ovarian endometrioid carcinomas also have simultaneous endometrial adenocarcinoma while 2-8% of patients with endometrial adenocarcinomas will have synchronous ovarian carcinoma; approximately 90% of synchronous tumors from ovary and endometrium will be endometrioid,

indicating they share common etiologic mechanisms.[165] Mabrouk *et al*., reported a case of mixed clear-cell and endometrioid type adenocarcinoma of the extragonadal, rectovaginal septum arising from endometriosis and associated with differentiated, endometrioid endometrial carcinoma.[166]

A premalignant lesion known as endometrial hyperplasia, is estimated to progress into invasive endometrial cancer in 5-10% of benign tumors and 20-30% of cases with accompanying atypia.[167] Endometrial hyperplasia exhibits a frequency of 15% of *K-RAS* mutations similar to that seen in endometrial adenocarcinoma suggesting that this may be an early event in endometrial carcinogenesis.[168] Maxwell *et al*., identified *PTEN* somatic mutations in 10/51 (20%) endometrial hyperplasia cases with no delineation in frequency between those with or without atypia, indicating that the mutation of *PTEN* is an early event in some endometrial cancers.[169] Moreover, patients with endometriosis exhibit endometrial hyperplasia and are at an increased risk of neoplastic transformation, though the exact risk of development into adenocarcinoma is currently unknown.[170] The simultaneous detection of endometriosis, endometrial, and ovarian cancers, genetic mutation semblance in *K-RAS* and *PTEN* for endometrial cancer and endometrial hyperplasia, as well as the association of hyperplasia and atypia in endometriosis suggests that endometriosis may be a neoplastic process of endometrial cancer.

1.3.3 Endometriosis and ovarian cancer risk

Of all neoplasms, ovarian cancer has been the most consistently associated with endometriosis.[171] This association was first identified in 1925[52,131] and has since been observed in a number of clinical cases among women with long-standing ovarian endometriosis[126] and has been shown to occur at a high rate, in 4-29% of all cases.[152,153] Epidemiologic, histopathologic, and molecular data suggest endometriosis may be a precursor lesion to specific types of ovarian cancer.[149] Based on molecular genetic studies, epithelial ovarian cancer can be divided into two

categories: **type I** (endometrioid, clear-cell, mucinous, micropapillary serous carcinomas, and low-grade serous carcinomas) and **type II** (high grade serous carcinomas, carcinosarcoma, and undifferentiated carcinomas).

Type I tumors develop slowly, and progress through a borderline precursor tumor stage such as endometriosis or low-malignant potential tumors. They are low-grade, confined to the ovary at diagnosis, genetically stable, and comprise ~25% of all ovarian cancers.[130] They also demonstrate mutations in *K-RAS*, *BRAF*, *CTNNB1*, *PTEN*, *TP53*, *BCL-2,* and *ARID1A* genes.[79,149,165,172] The greatest risk of endometriosis-associated malignancy are endometrioid and clear-cell carcinomas.[173] Type II tumors are rapid growing, aggressive, malignant tumors without known precursor tumors. They are high-grade, with an origin of tubal epithelium, genetically unstable and demonstrate mutations in *PIK3CA* and *TP53* genes,[79,149,172] and *HER2* expression.[174] Currently no association has been reported between endometriosis and mucinous or high-grade serous ovarian cancer.[173]

Novel insights of the malignant transformation from endometriosis into ovarian carcinoma have identified pathologically, a morphological continuum of sequential steps from normal endometriotic epithelium to atypical endometriosis and finally to invasive ovarian carcinoma.[175] In this study, researchers identified that atypical endometriosis and endometriosis-related ovarian neoplasms share molecular alterations such as *PTEN* mutations, with ovarian endometrioid carcinoma harboring mutations in *PTEN* in 14-20% of the cases.[175] This underpins a previous study where PTEN somatic mutations were identified in 4/20 ovarian endometrioid carcinomas (20%), 2/24 clear-cell (8.3%), and 7/34 solitary endometriotic cysts (20.6%).[176] In fact, 5/7 of the endometrioid cases and 7/24 clear-cell cases had synchronous endometriosis. A separate study identified 22/37 of ovarian carcinomas with endometriosis demonstrated transition from typical endometriosis to atypical endometriosis, while 23/37 cases demonstrated atypical

endometriosis to carcinoma.[177] This supports the emerging theory that endometriotic lesions potentially harbor genetic defects that are permissive to malignant transformation.

1.3.4 PTEN somatic missense mutations associated with endometriosis, endometrial cancer, and ovarian cancer

The tumor suppressor function of PTEN is dependent on its phospholipid phosphatase activity and the loss-of-function of the phosphatase catalytic domain is commonly associated with oncogenic *PTEN* mutations.[4,178-180] Recently somatic mutations and deletions of *PTEN* have been reported in many types of sporadic tumors,[181] including endometriosis,[147] ovarian, and endometrial cancers.[147,156,176,182-190] The incidence of *PTEN* mutations in endometrial tissue of women diagnosed with endometriosis and endometrial hyperplasia is one of the highest among analyzed tumors and the most commonly mutated gene identified in endometrial cancer.[169,182] In fact, data suggests that *PTEN* is more commonly mutated than any other gene including *K-ras* and *p53* in ovarian and endometrial cancers.[169,182] The inactivation of *PTEN* has been described as an early event in endometrial hyperplasia and the development of ovarian and endometrial cancers.[169,176,191] Thus, mutations and/or the loss of *PTEN* may contribute to the genesis and development of endometriosis and subsequently cancer. Therefore a thorough structural investigation of *PTEN* somatic mutations associated with endometriosis and endometrial and ovarian cancers will not only advance the understanding of genotype-phenotype interactions, but also further illuminate its molecular etiology.

1.4 Hypothesis and scope of research

Insights into the structural properties of PTEN provide a platform to analyze the mutations found in endometriosis, endometrial cancer, and ovarian cancer. Recent computational studies on PTEN reveal that missense mutations affect its function and structure.[192,193] Therefore, a careful dissection into the effects of putative mutations is pertinent to understanding the molecular mechanism in each of the identified phenotypes proposed herein.

In this study, thirteen somatic missense mutations were identified from published literature, offering a starting point to explore genotype-phenotype correlations in endometriosis, endometrial cancer, and ovarian cancer. We posit that the active-site of PTEN defines a possible mutation-driven allosteric region wherein a small subset of mutations correlate with endometriosis, endometrial cancer, and ovarian cancer. Our study focuses on understanding the impact that these missense mutations have on the structure of PTEN and on investigating the molecular mechanism of each of the phenotypes. It is our hope that the results from our study will aid in a better clinical-molecular classification of the resulting phenotypes and allow translation into improved diagnostic and therapeutic approaches.

Chapter 2: Theoretical background and methods

2.1 pKa prediction

The structure, stability, and function of proteins are dependent on their net charge and on the ionization state of individual residues.[194] In essentially all biological reactions, there exists at least one step that involves the transfer of one proton group to another.[195] The ease at which this transfer takes place (between an acid and a base) depends on the relative proton affinities of each group.[195] Proton affinity is measured by the pKa value of an ionizable residue and is the logarithmic measure of the acid dissociation constant of a titratable group, which can be utilized as an indicator to determine the ionization of an amino acid at a given pH. The quantitative relationship between the pH of the solution and ion concentration is represented by the following ionization equilibrium reaction and Henderson-Hasselbach equation:

$$HA + H_2O \rightarrow A^- + H_3O^+$$

(Equation 2.1 *Ionization equilibrium reaction*)

$$Ka = \frac{[A^-][H_3O^+]}{[HA]}$$

$$pH = pKa + \log\frac{[A^-]}{[HA]}$$

(Equation 2.2 *Henderson-Hasselbach equation*)

The pKa of a titratable site is therefore equal to the pH at which the site is 50% occupied by a proton. **Table 2.1** references pKa's of common ionizable residues alone in solution. The pKa values of ionizable residues however encounter several differences inside folded proteins compared to alone in solution. The factors that affect perturbation are a result of interactions of charged ionizable groups with neighboring residues (site-site interactions), electrostatic permanent dipoles in the protein (background interactions), as well as the altered interactions with

water (desolvation energy).[194,196] However, perturbation is mostly due to electrostatic interactions of nearby residues.[197]

Table 2.1 pKa of ionizable amino acids in solution

Ionized state	Neutral state	pKa (solution)
N-terminus, *α-Amino* ($-NH_3^+$)	R - NH_2	9.8
C-terminus, *α-Carboxyl* ($-COO^-$)	R - COOH	2.2
***Guanidino*, Arg ($-NH_2$+)**	Arg (-NH)	12.4
***ε-Amino*, Lys ($-NH_3^+$)**	Lys ($-NH_2$)	10.4
***Phenol*, Tyr ($-O^-$)**	Tyr (-OH)	10.0
***Sulfhydryl*, Cys ($-S^-$)**	Cys (-SH)	8.3
***Imidazole*, His ($-NH^+$)**	His (-N)	6.3
Non-*α-Carboxyl,* Glu ($-COO^-$)	Glu (-COOH)	4.4
Non-*α-Carboxyl,* Asp ($-COO^-$)	Asp (-COOH)	4.0

Knowing the pKa's of residues and associated pKa shifts in their proteinaceous state allows for the determination and relative importance of the factors that affect perturbation in the aforementioned. These pKa shifts reveal how much an ionizable residue retains or releases a proton and provides insight in to the local environment as to whether a residue is involved in catalysis, solvent exposed, or buried. Moreover, it is important to know that the charges and dipoles of the protein may stabilize or destabilize a charged state in the protein.[198] Likewise, the desolvation impact when a charged group is brought from the highly polar aqueous solution into the typically apolar interior of the protein destabilizes charged effects.[198] These are important factors to consider since when proteins are folded pKa perturbations on the surface of the protein are typically small and are determined by charge-charge interactions with other ionizable groups.[199]

However, if these groups are partially or fully buried in the protein interior, large positive and negative perturbations often occur.[200] Moreover, experimentally, mutations of charged

residues in proteins have been shown to be significant for the determination of molecular conformation.[201] The suitable place to start when considering the calculation of pKa's of each ionizable residue is to consider the difference between the pKa value of a model in its aqueous state as well as the pKa value of the ionizable residues in the protein. The pKa predictions evaluate the electrostatic potentials of ionizable residues by solving the Poisson-Boltzmann (PB) equation in order to produce protonation equilibria in proteins.[197]

The University of Houston Brownian Dynamics (UHBD) program[202] provides fast evaluation of various electrostatic energy terms for all ionizable residues of a protein utilizing the finite-difference Poisson-Boltzmann method (FDPB). This method is a numerical solution to the linearized Poisson-Boltzmann equation by which to calculate the self and interaction energies[202-204] of the ionizable groups in the protein:

$$\nabla \cdot \varepsilon(r)\nabla\phi(r) = -4\pi\rho^{f}(r) + \lambda(r)\bar{\kappa}^{2}\phi(r) \qquad \text{(Equation 2.3 } \textit{Poisson-Boltzmann}\text{)}$$

where $\boldsymbol{\varepsilon}$ is the dielectric constant, $\boldsymbol{\phi}$ is the electrostatic potential at position $\boldsymbol{r}$ to the density of the charge distribution $\boldsymbol{\rho^{f}}$, $\boldsymbol{\lambda}$ delineates regions accessible to mobile ions and $\boldsymbol{\overline{\kappa}}$ is a modified Debye-Hückel parameter. The FDPB method allows the calculation of the free energy of a macromolecule in a given ionization state. These energies are required for the pairwise-interaction energy grid utilized to calculate ionizable residue pKa shifts within a system. The free energy change of a given residue from the deprotonated state to the protonated state is obtained in the following equation:

$$\Delta G = 2.303\, RT\gamma\left(pH - pK_a^i\right) \qquad \text{(Equation 2.4 } \textit{Free energy change}\text{)}$$

where $\boldsymbol{pK_a^i}$ is the intrinsic (experimental) pKa of the residue in solution and **pH** is that of the

environment. Thus the free energy of the ionization state ($x_1,\ldots,x_M$) of the entire protein is:

$$\Delta G(x_1, \ldots, x_M) = 2.303RT \sum_i x_i y_i (pH - pK_{i,protein})$$

(Equation 2.5 *Free energy of ionization state*)

where $\boldsymbol{x_i}$ is 1 when the group $\boldsymbol{i}$ is ionized and 0 when it is neutral. This equation allows for the analysis of electrostatic interaction of titratable residues within a protein, thus for a protein with M ionizable groups, the M ionization states and free energy state is expressed as:

$$\Delta G(pH, x_1', \ldots, x_M') = 2.303RT \sum_{i=1}^{M} x_i' (pH - pK_{i,instrinsic}) + \sum_{i=1}^{M} x_i' \gamma_i G_{ij} + \sum_{i=1}^{M} \sum_{j=i+1}^{M} [(q_i^\circ + x_i')(q_i^\circ + x_j') - q_j^\circ q_j^\circ] \psi_{ij}$$

(Equation 2.6 *Additivity of electrostatic potentials*)

2.2 Generation of PTEN mutants

In silico modeling requires the three-dimensional structure of nucleic acids, proteins or biomolecular complex of interest. The Protein Data Bank (PDB)[205] is a database that contains the structural data of biological molecules that were either determined via X-ray diffraction or through NMR experiments with approximations that comprise the atomic coordinates of the molecule of interest. Alternatively, if a structure has not been determined, a technique that enables the construction of a protein structure with known sequence can be carried out utilizing homology modeling which compares a protein with known sequence with an experimentally determined structure.[206] If a mutant structure is required yet does not exist, one can be generated

in silico utilizing a variety of programs. One commonly used program is the Mutator module (v1.3) within the Visual Molecular Dynamics program (VMD).[207] Mutant structures are built *in silico* by side-chain replacement starting from a predetermined wild-type (WT) protein structure.

2.3 Electrostatic surface potential (ESP) calculation

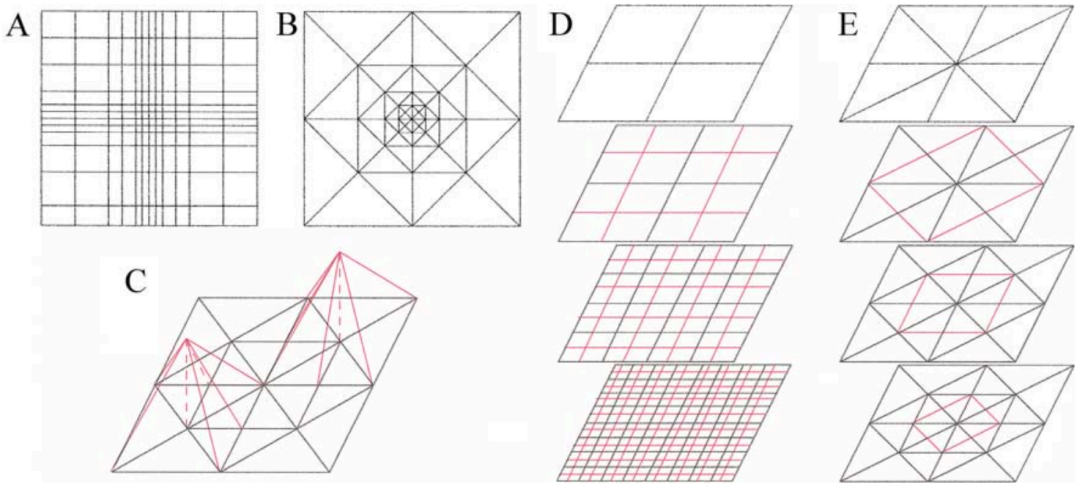
Electrostatic interactions are a crucial property to understand as these interactions govern the folding of many proteins. A strong electrostatic field is produced by the many biological macromolecules that contain charged groups that are close to each other or at the surface and contribute to the overall charge density. Indeed, such interactions are used to stabilize a protein structure and are essential in governing protein-protein and protein-ligand interactions. These long-ranged interactions can accelerate the rate at which molecules associate.[208] Electrostatic forces can steer the ligand into its binding site on the protein, which is critical to the function of the protein. In fact, weakly polar interactions depend on the electrostatic attraction between opposite charges. Some binding sites are naturally shielded from the solvent and are "closed" by salt links between groups on the protein surface. If the correct 'charged' substrate disrupts these salt links, it gains access to the binding site. Additionally, Electrostatic interactions between ligand and protein can contribute to the affinity and specificity of binding and orientation of the ligand. It is therefore essential to assess the surface charge distribution of a protein by calculating the electrostatic surface potential (ESP).

Upon completion of the pKa calculation of each ionizable residue, an electrostatic surface potential map can be generated. In calculating electrostatics, one can solve the linear and non-linear Poisson-Boltzmann (PB) equation using a finite difference method. The PB equation is a fundamental equation of classical electrostatics and is defined as:

$$\nabla \cdot \varepsilon(r)\nabla\varphi(r) = \rho(r) \qquad \text{(Equation 2.7 } \textit{PB equation}\text{)}$$

where $\boldsymbol{\varphi}$ is the electrostatic potential, $\boldsymbol{\varepsilon}$ is the dielectric constant (relative permittivity) of the medium (can be different throughout different regions of space), $\boldsymbol{\rho}$ is the charge density due to atomic charges and $\mathbf{r}$ is the position. $\boldsymbol{\varphi}$, $\boldsymbol{\varepsilon}$, and $\boldsymbol{\rho}$ are all dependent on position. The Poisson equation indicates how the electrostatic potential due to distribution of charges in a heterogeneous dielectric environment varies throughout space.[208] The Boltzmann equation indicates how the distribution of charged atoms of the salt varies with the electrostatic potential. Both of these equations must be solved simultaneously in order to determine the actual potential.

Figure 2.1 Discretization schemes and hierarchies utilized in Poisson-Boltzmann solvers



(A) Cartesian mesh suitable for finite difference (FD) calculation, (B) Finite element (FE) mesh exhibiting adaptive refinement, (C) Piecewise linear basis function used in FE methods, (D) Multi-level hierarchy used to solve the PB equations for an FD, (E) Multi-level hierarchy used to solve the PB equation for FE discretizations; red lines denote the simplex subdivisions used to introduce additional unknowns at each hierarchy level.

From Baker, N.A.[209]

Methods for solving the PB equation have been developed to improve the efficiency of electrostatic calculations the most common methods include: finite difference (FD), finite element

(FE), and boundary element methods (BEM) (see **Figure 2.1**) - the former being the most popular method.[209] In the **finite difference (FD) method,** the space is discretized (chopped up in discrete chunks).[208] The FD method solves the PB equation on a non-uniform Cartesian mesh for a two-dimensional domain. The Poisson differential operator is transformed in a sparse difference matrix by means of a Taylor Series expansion. The resulting matrix equations are then solved by various linear algebra techniques. In the adaptive **finite element (FE) method** finite element meshes are composed of triangles or tetrahedra that are joined at edges and vertices. Solution accuracy can be increased in specific areas by locally increasing the number of vertices through simplex refinement.[209] The ability to locally increase the solution resolution is called "adaptivity" and is the major strength of FE methods applied to the equation.

2.4 Multiple sequence alignment (MSA) – sequence conservation

Multiple sequence alignment (MSA) involves the alignment of three or more biological sequences of similar length. From the output, homology can be inferred and the evolutionary relationships between the sequences are assessed. There are multiple methodologies to compute MSAs some of which require more sophisticated methodologies than **pairwise alignment.** However, most MSA programs utilize **heuristic (progressive) methods**, as it can be a prohibitive computational expense in calculating an optimal alignment. The most widely utilized progressive alignment algorithm is CLUSTAL W.[210] Three steps utilized in all progressive alignment algorithms are: (i) calculation of pairwise distances based on pairwise alignment of the sequences (percent identity), (ii) build a guide tree which is an inferred phylogeny for the sequences, and (iii) utilize the tree to guide the progressive alignment. There are many ways to build a tree from a matrix of pairwise distances. CLUSTAL W uses the neighbor-joining method.[210] A root of the

tree is then determined by the mid-point method, which gives equal means for the branch lengths on either side of the root. The **W**eights (CLUSTAL **W**) are then calculated correcting for unequal sampling at different evolutionary distances.[210] A multiple alignment editor such as Jalview[211] can be utilized to align and visualize the sequences.

2.5 Elastic network model – normal mode analysis (NMA)

Normal mode analysis (NMA) is a simulation technique utilized to probe, large scale motions in macromolecular molecules. Traditionally used in infrared and Raman spectroscopy experimental techniques, NMA has demonstrated in more recent developments, that it is capable of providing unique insights into the structural and dynamic properties of complex systems as well as in the prediction of functional motions in biomolecules. Formally NMA involves a Hessian matrix inversion process that limits its applicability to large systems; however both numerical and approximate methods have been developed to extend the size limit. This can be computational demanding as it involves the diagonalization of a $3N$ x $3N$ matrix, where N is the number of atoms in the molecule.

One of the more recent developments is the use of simplified coarse-grained (CG) NMA, known as **elastic network model (ENM)** NMA as an alternative to full atomic, "standard" NMA. This method has two advantages over the standard NMA: (i) energy minimization is not necessary because all of the distances of all of the elastic connections are taken to be at their minimum-energy length (ii) diagonalization is substantially reduced because the number of atoms is reduced as only Cα atoms are utilized. This results in a ten-fold reduction in the number of atoms, thus greatly reduces the computational time. In contrast to standard NMA, ENMs require two assigned parameters: (i) the force or spring constant denoted as γ, and (ii) a cut-off distance denoted as, R_c.[212] As evidenced by a pioneering study of Tirion,[213] a hypothetical force field with

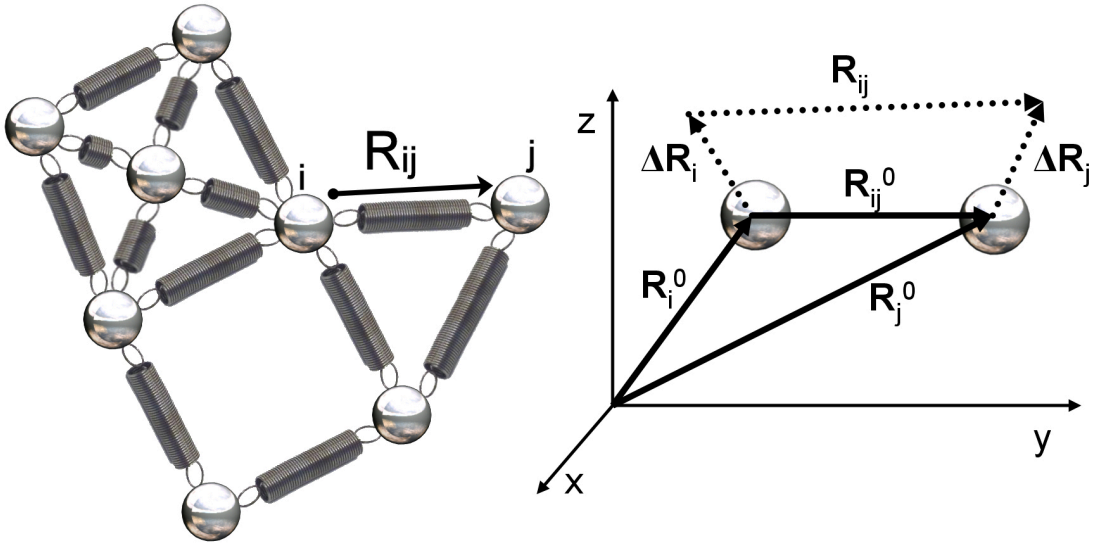
uniform (single-parameter) harmonic potentials yields global modes almost indistinguishable from those obtained from a detailed force field with specific nonlinear terms.[214] The shape of the global modes is dominated by the network of inter-residue contacts, which is a purely geometric quantity defined by the overall shape, form, or native contact topology of the protein.[214]

Inspired by the work of Tirion,[213] the **Gaussian network model (GNM)**[215,216] was the first simplified ENM and is based on the theory of elasticity.[217-221] The GNM involves a single site per residue representation, in which the sites are identified by the Cα atoms and described as an elastic network connected by a uniform spring force constant, **γ,** provided that they are within a certain cut-off distance, R_c (**Figure 2.2**). The dynamics of the interconnected bead-and-spring model best described by the GNM, uses the following potential:

$$V_{GNM} = \frac{\gamma}{2}\sum_{i,j}^{N}\left(\boldsymbol{R}_{ij} - \boldsymbol{R}_{ij}^{0}\right)^{2} H\left(r_c - r_{ij}\right) \qquad \text{(Equation 2.8 } \textit{GNM potential}\text{)}$$

where **γ** is the uniform spring constant, $\boldsymbol{R}_{ij}^{0}$ and $\boldsymbol{R}_{ij}$ are the original and instantaneous distance vectors between *i* and *j*; the summation is performed over the pairs of residues/nodes filtered through the heavy step function $H\left(r_c - r_{ij}\right)$ and selects the interacting pairs within the cut-off of r_c and is also equal to 1 if the argument is positive, and zero otherwise.

Figure 2.2 Harmonic spring network interaction sites and displacements



From https://en.wikipedia.org/wiki/Gaussian_network_model

The GNM is based on the assumption that all residue fluctuations (and inter-residue distances) are Gaussianly distributed around their equilibrium coordinates. The equilibrium coordinates are identified by the position vectors $\boldsymbol{R}_{ij}^{0}$ of Cα atoms in a protein structure. The Hessian matrix (standard NMA) is replaced by an *N x N* Kirchoff matrix, which describes the inter-residue contact topology, **Γ,** such that *N*-1 isotropic modes are obtained, and is the sole determinant of equilibrium dynamics. The fluctuations and their cross-correlations are fully controlled by the matric. The Kirchoff matrix, **Γ** is defined by:

$$\Gamma_{ij} = \begin{cases} -1 & if\ i \neq j\ and\ R_{ij} < r_c \\ 0 & if\ i \neq j\ and\ R_{ij} > r_c \\ -\sum_{j,j\neq 1}^{N} \Gamma_{ij} & if\ i = j \end{cases}$$

(Equation 2.9 *Kirchoff matrix*)

where $\mathbf{r_c}$ is a cutoff distance for spatial interactions, usually 7 Å for proteins.[222]

Another broadly utilized ENM is the **anisotropic network model (ANM)**.[223-226] Similar to the GNM, the positions of the nodes in the ANM are identified but the Cα atoms, however, the ANM potential is defined as:

$$V_{ANM} = \frac{\gamma}{2}\sum_{i,j}^{N}\left(R_{ij} - R_{ij}^{0}\right)^{2} H\left(r_c - r_{ij}\right) \qquad \text{(Equation 2.10 } \textit{ANM potential}\text{)}$$

with the central difference between GNM and the above potential being that the vectors $\boldsymbol{R}_{ij}^{0}$ and $\boldsymbol{R}_{ij}$ in **equation 2.10** replaces distances (scalars), R_{ij}^{0} and R_{ij}. This indicates that the potential depends on their scalar product rather than upon the dot product between the fluctuation vectors as in the GNM. Here $V = 0$, if $R_{ij} = R_{ij}^{0}$ irrespective of the direction of the corresponding distance vectors. An $\boldsymbol{r_c}$ = 12-15 Å provides a good description of the native contact topology. The ANM requires the eigenvalue decomposition of the Hessian $\boldsymbol{H}$ (second derivative of potential, $\boldsymbol{V}$), a 3N x 3N matrix composed of N x N elements, $\boldsymbol{H}_{ij}$ $(i \neq j)$ defined as:

$$\boldsymbol{H}_{ij} = \frac{\gamma\Gamma_{ij}}{\left(R_{ij}^{0}\right)^{2}} \begin{matrix} X_{ij}X_{ij} & X_{ij}Y_{ij} & X_{ij}Z_{ij} \\ Y_{ij}X_{ij} & Y_{ij}Y_{ij} & Y_{ij}Z_{ij} \\ Z_{ij}X_{ij} & Z_{ij}Y_{ij} & Z_{ij}Z_{ij} \end{matrix} \qquad \text{(Equation 2.11 } \textit{Hessian matrix}\text{)}$$

where $X_{ij}, Y_{ij},$ and Z_{ij} are the components of the distance vector, R_{ij}^{0} . The pseudo-inverse of $\boldsymbol{H}$ is the 3N x 3N covariance matrix, C^{ANM}, which can be expressed in terms of the 3N-6 non-zero eigenvalues $\boldsymbol{\lambda}_k$ and corresponding eigenvectors $\mathbf{u}_k$ of $\boldsymbol{H}$ and described as:

$$C^{ANM} = \sum_{k=1}^{3N-6}\frac{1}{\lambda_k}u_k u_k^{T} \qquad \text{(Equation 2.12 } \textit{ANM covariance matrix}\text{)}$$

The ANM has been shown alongside standard NMA to match closely the collective

motions along the second lowest frequency mode intrinsically accessible to the original structure of hemoglobin demonstrating the allosteric change in conformation from its tense form (T) to its relaxed (R) form.[227,228] The main advantages of ENM-based approaches such as the GNM and the ANM are for their (i) ability to provide an exact solution for the unique dynamics of a structure and (ii) applicability to large biomolecular complexes and assemblies beyond the range of atomic simulations.[214] It is suggested that the GNM be utilized (rather than ANM) for evaluating mean-square fluctuations of residues and their cross-correlations, as well as the profiles of residue displacement sizes in collective modes, and that the ANM be utilized for exploring properties that cannot be calculated using the GNM such as deformation vectors or molecular motions.[229]

2.6 Elastic network model – protein structure network (PSN)

Protein structure network (PSN) analysis is an extension of the ENM-NMA outlined in the previous section see (section 2.5). The concept of PSN has been explored more recently as it offers insights into the global properties of protein structures.[230,231] The representation of protein structures as networks of interactions between amino acids has proven useful in a number of studies, such as protein folding[232] and prediction of functionally important residues in enzyme families;[233] therefore contributing to the issue of intra-molecular and inter-molecular communication.[230,231] Based on early work established by Vishveshwara *et al.*,[234,235] PSN is constructed from the atomic coordinates of residues, which represent the nodes of the network. Two nodes are connected by an edge if the percentage of the interaction between them is greater than or equal to a given Interaction Strength cut-off.

$$I_{ij} = \frac{n_{ij}}{\sqrt{N_{ij}N_{ij}}}100 \qquad \text{(Equation 2.13 } \textit{PSN-Interaction Strength cut-off}\text{)}$$

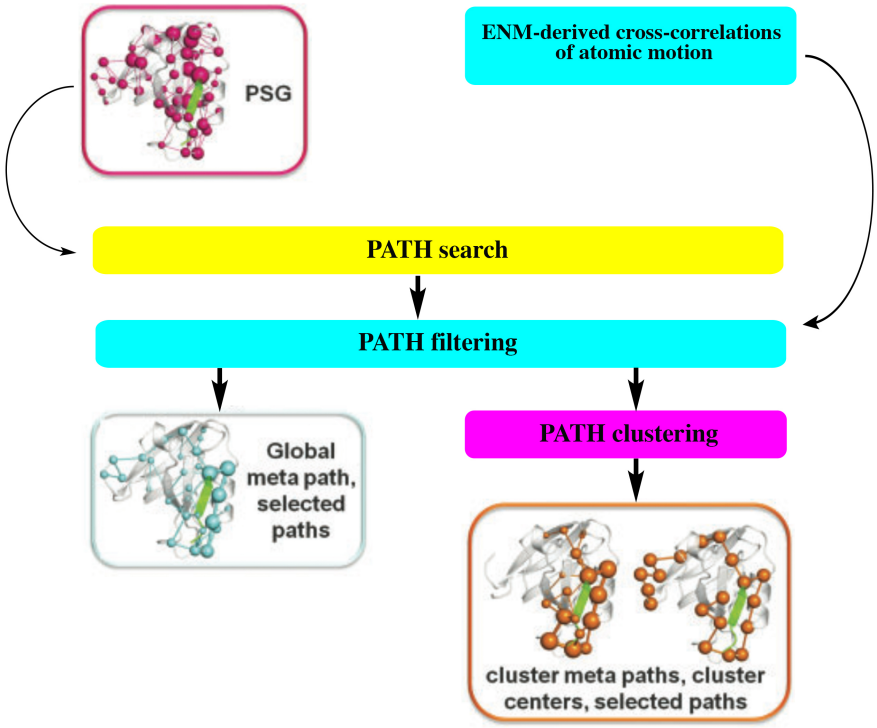
where $\boldsymbol{I_{ij}}$ is the interaction percentage of nodes $\boldsymbol{i}$ and $\boldsymbol{j}$, is the number of side-chain atom pairs

within a given cut-off (4.5 Å), and N_i and N_j are, respectively, the normalization factors (NF) for residues ***i*** and ***j***, which take into account the difference in size of different nodes and their propensity to make the maximum number of contacts with other nodes within protein structures.

This PSN method lends itself to a recently implemented mixed PSN-ENM approach to predict salient structural communication pathway in biomolecular systems. The underpinning of WebPSN[236] lies with a Protein Structure Graph (PSG) and searches for all shortest communication pathways between user-specified residues. A PSG defines amino acids as nodes and the noncovalent interactions among them as links. Such graphs are useful in identifying clusters of residues that stabilize the protein structure and protein-protein interfaces.[234] The network topology of PSGs depends on the cut-off of the interaction strength between residues used in the constructed graph.

The mixed PSN-ENM method (WebPSN) involves a multi-step process (**Figure 2.3**) where network features (i.e. nodes, hubs, links) are computed by building a PSG and the shortest communication pathways on ensembles of structures are acquired from a single high-resolution structure. The algorithm defines all possible communication paths between selected node pairs and filters the results to cross-correlation of atomic motions, as derived from ENM-NM. Filtering consists in retaining only the shortest path(s) that contains at least one residue correlated (i.e. with a cross-correlation value ≥0.6) with either one of the two extremes (i.e. the first and last residues in the path). Meta-paths made of the most recurrent nodes and links in the path pool (i.e. global meta-paths) and infer a coarse/global picture of the structural communication in the considered system. In detail, meta-paths are made of nodes ≥5% of the considered path pool (i.e. 'frequent nodes'), and of links satisfying both conditions of being present in of the paths and of connecting 'frequent nodes'.

Figure 2.3 Mixed PSN-ENM WebPSN flowchart



Adapted from Seeber, M. *et al*.[236]

A recent study demonstrated the advantages of utilizing this mixed PSN-ENM approach on a domain in its free state and in complex with a C-terminal peptide illustrating salient communication features within each system.[237] The benefits of this high-speed approach are well-suited for high-throughput investigation of the structural communication pathways including allosterism, in large biomolecular systems in different functional states.[236]

2.7 Structural stability prediction

Many computational methods have been developed to predict the overall structural stability of a protein as the result of the introduction of a mutation. The core functionality of these methods involves an energy function that calculates the free energy of the system. Many of these functions were also developed to estimate the consequence of a mutation on protein stability or binding affinity.[238] Though three-dimensional structures provide atomic details on their architecture, but not on the forces stabilizing them.[238] Introducing mutations and quantifying their energetic consequence can assist in assessing these forces. Most of the mutations either are neutral or destabilize protein structure; however many stabilizing mutants have demonstrated to be advantageous to protein function.[239]

Three common programs utilized to predict protein structure stability are: I-Mutant 3.0,[240] CUPSAT,[241] and Elastic Network Contact Model (ENCoM).[242,243] **I-Mutant 3.0** is a support vector machine (SVM)-based tool that calculates protein stability changes upon mutation either from sequence or structural information.[240] The structure-based version of I-Mutant 3.0, classifies the prediction into one of three classes: neutral mutation ($-0.5 \leq \Delta\Delta G \geq 0.5$ kcal/mol), large decrease (<-0.5 kcal/mol), or large increase (>0.5 kcal/mol).[240] The predicted free energy change $\Delta\Delta G$ is based on the difference between unfolding Gibbs free energy change of mutant and native protein (kcal/mol).[240] **CUPSAT** is a physiochemical modeling program that utilizes a Boltzmann mean-force potential and torsion angles (phi, psi) to predict $\Delta\Delta G$ values and stability changes.[241] CUPSAT also calculates the overall stability change as well as the adaptation (favorable or unfavorable) of the observed torsion angles (phi, psi) as a result of the mutation. **ENCoM**[242,243] utilizes a novel mixed coarse-grained normal mode method to account for the type and extent of pairwise atomic interactions allowing for the calculation of vibrational entropy differences as a result of mutations.[242,243] ENCoM utilizes a potential function with four terms: (1) covalent bond

stretching, (2) angle bending, (3) dihedral torsion, and (4) non-bonded interaction while taking into consideration the nature and possible effects that the orientation of the side-chains have on dynamics within the context of normal mode analysis. Utilizing computational methods to predict structure stability changes upon mutation can assist in designing new or mutated protein with specific levels of stability, enzymatic activity and binding to other molecules (proteins, DNA, ligands, substrates, etc.) further stabilizing its function.[238]

2.8 Molecular dynamics (MD) simulations

In modern molecular modeling, molecular dynamics (MD) simulation is a very powerful technique that aims to understand the properties of assemblies of molecules in terms of structure and microscopic interactions between them by simulating their behavior *in silico*.[244] Molecular dynamics enables the study of complex dynamic processes that occur in biological systems and is a useful technique for the characterization of macromolecular stability, protein folding, dynamics and flexibility, conformational and allosteric properties, molecular recognition (properties of complexes), and ion transport on the nanosecond, microsecond and millisecond time scale.[245,246] Extraction of these macroscopic observables, therefore, allows for the prediction of mechanisms and a greater understanding of conformational changes in a protein.[247] Molecular dynamics offers the opportunity to perform a wide variety of studies, which include both drug and protein design.

Molecular dynamics simulations are in many respects very similar to conventional (wet-lab) experiments. When a conventional experiment is performed, the following occurs: preparation of sample, connect sample to instrument (i.e. thermometer, fluorescence microscope), and measure the property of interest over a certain interval.[248] In a molecular dynamics simulation, the same approach is followed: first a sample is prepared by selecting a model system of N particles and then Newtown's equations of motion are solved until the properties of the

system no longer change over time (equilibration). After equilibration, the actual measurement is performed as a function of the positions and momenta of the particles in the system.[248] These simulations are referred to as entirely classical since only classical Newtonian dynamics (equations of motion) are utilized – quantum and reactive processes are not calculated.

As conventional experimental approaches reach a maximal resolution of ms-μs in time, and a length order of nm, theoretical (simulation) models allow simulating the microscopic level at atomic resolution, but only on the order of 10^2 ns/10^1 nm for small systems. For most biomolecular systems, however, it is often necessary to work with empirical parameterizations of models, which are orders of magnitude faster and have the ability to produce observations on a much greater timescale (**Table 2.2**).

Table 2.2 Timescales and amplitudes of biomolecular motions

Timescale	Amplitude	Description
		Local Motions
Short: femto, pico $\mathbf{10^{-15} – 10^{-12}}$ **s**	0.001 – 0.1 Å	-Atomic fluctuations: bond stretching, angle bending, dihedral motion -Side-chain motions -Loop motions
		Rigid Body Motions
Medium: pico, nano $\mathbf{10^{-12} – 10^{-9}}$ **s**	0.1 – 10 Å	-Helix motions -Water relaxation -Collective motions: subunit ad domain motions (hinge bending, twisting)
		Large-Scale Motions
Long: nano, micro $\mathbf{10^{-9} – 10^{-6}}$ **s**	1 – 100 Å	-Folding in small peptides -Helix coil transitions -Dissociation/association -Ion transport
		Very Large-Scale Motions
Very long: micro, second $\mathbf{10^{-6} – 10^{-1}}$ **s**	10 – 100 Å	-Folding and unfolding -Ribosome synthesis

Essentially, molecular dynamics generates configurations of a N-body system by integration of Newton's laws of motion and calculates the time-dependence of that molecular system.[249,250] To begin, a MD simulation algorithm needs only an initial configuration of the system, *r* (atomic coordinates), and a force-field, *E(r)*. Given the potential and force (negative gradient potential) acceleration can be calculated:

$$\overrightarrow{f_i} = \frac{\partial E_i}{\partial \vec{r}_i} = m_i \vec{a}_i \qquad \text{(Equation 2.14 } \textit{Force}\text{)}$$

From the integration of the acceleration, velocities are calculated:

$$\overrightarrow{v_i} = \int \overrightarrow{a_i}\, dt \qquad \text{(Equation 2.15 } \textit{Velocities}\text{)}$$

With the integration of velocities, a new set of positions are calculated:

$$\overrightarrow{r_i} = \int \vec{v}_i \; dt \qquad \text{(Equation 2.16 } \textit{Positions}\text{)}$$

The equations of motion are integrated numerically and broken into many small fixed time steps, δt.[250] The total force on each atom in the configuration is calculated from the vector sum of interactions with other atoms at time *t*. Using the Newtonian equation, an algorithm iteratively calculates forces by solving equations of motion and assigning positions based on accelerations, with increasing time. The resulting trajectories are defined by both position and velocity vectors and describe the time evolution of the system representing the conformational phase space of the system.

2.8.1 The Verlet and leap-frog algorithms

Numerous algorithms exist for integrating the equations of motion and can be a rather simple calculation for a two-atom system. However, the number of atoms increases for a system with many particles, for which the finite difference approach is used to integrate the equations of

motion. The algorithm must be able to conserve energy and momentum, be time-reversible, and permit a long time step, δt. Given the molecular positions, velocities, and other dynamic information at time t, the positions and velocities can be obtained based on acceleration over time $(t + \delta t)$.[251] This loop of calculations is integrated iteratively at new positions, for a fixed time step in order to obtain new values at time $t + 2\delta t$.[250] All algorithms assume that the positions and dynamic properties (velocities, accelerations, etc.) can be approximated as Taylor series expansions:

$$r(t + \delta t) = r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t) + 1/6\ \delta t^3 b(t) + \ldots$$

$$v(t + \delta t) = v(t) + \delta t a(t) + \frac{1}{2}\delta t^2 b(t) + \ldots$$

$$a(t + \delta t) = a(t) + \delta t b(t) + \ldots \qquad \text{(Equation 2.17 } \textit{Taylor Series Expansions}\text{)}$$

$$b(t + \delta t) = b(t) + \ldots$$

From Leach, A.R., 2001[250] and Allen, M.P. and Tildesley, D.J.[251]

where $\boldsymbol{r}$ and $\boldsymbol{v}$ represent the complete set of positions and velocities (the first derivative of the positions with respect to time), $\boldsymbol{a}$ represents all accelerations (the second derivative), $\boldsymbol{b}$ denotes all third derivatives of $\boldsymbol{r}$.

The initial step in solving the equation of motion is to discretize the time by solving r(t) on a series of time instances t_i. The most employed algorithm in discrete time is the **Verlet**[252] algorithm, which utilizes positions and accelerations at time, t, as well as positions at the previous step r($t + \delta t$), to calculate the new positions at $t + \delta t$, r($t + \delta t$)[250] is given by:

$$r(t + \delta t) = 2r(t) - r(t + \delta t) + \delta t^2 a(t) \qquad \text{(Equation 2.18 } \textit{Verlet}\text{)}$$

A variation on the Verlet algorithm is the **leap-frog**[253] algorithm which calculates positions at time $\boldsymbol{t}$, velocities at time $\left(\boldsymbol{t} - \frac{\boldsymbol{1}}{\boldsymbol{2}}\boldsymbol{\delta t}\right)$ and is given by:

$$r(t + \delta t) = r(t) + \delta t v\left(t + \frac{1}{2}\delta t\right) \qquad \text{(Equation 2.19 } \textit{Leap-frog}\text{)}$$

$$v\left(t + \frac{1}{2}\delta t\right) = v\left(t - \frac{1}{2}\delta t\right) + \delta t a(t)$$

From Allen, M.P. and Tildesley, D.J.[251]

Here the velocities are always calculated at intermediate steps between two subsequent position calculations. The stored quantities are the current positions *r(*t) and accelerations *a*(*t*) together with mid-step velocities $v\left(t + \frac{1}{2}\delta t\right)$. The velocity equation is implemented first and the velocities leap over the coordinates to give the mid-step values. This is necessary so that the energy at time *t* can be calculated as well as any other quantities that require positions and velocities at the same instant.[251]

2.8.2 Potential energy function

In addition to the coordinates and velocities, the forces of the system are calculated by an empirical potential energy function, or force field. The effective molecular energy can be described as a sum of potentials derived from simple physical forces, which include both bonded (covalently linked atoms) energy terms, non-bonded (non-covalent) interaction terms and optional terms (constraints):

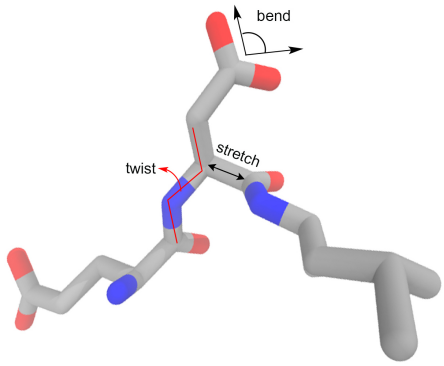
$$E_{Total} = E_{bonded} + E_{nonbonded} + E_{optional\ terms}$$

(Equation 2.20 *Potential energy, Force field*)

For bonded (intramolecular) interactions the potential is usually a sum of harmonic and Fourier terms in order to model energy of stretching, bending, and torsion motions between covalently bound atoms described as:

$$E_{bonded} = E_{bonds} + E_{angles} + E_{dihedrals}$$

Figure 2.4 Potential energy function (bonded terms)



$$= \Sigma_{bonds} \frac{k_i}{2} (l_i - l_{i,0})^2 + \Sigma_{angles} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 + \Sigma_{\substack{improper \\ dihedrals}} \frac{k_i}{2} (\xi_i - \xi_{i,0})$$

$$+ \Sigma_{\substack{proper \\ dihedral}} \frac{V_n}{2} (1 + \cos(n\omega - \gamma))$$

(Equation 2.21 *Bonded terms*)

where $\boldsymbol{l}$ represents bond length, $\boldsymbol{\theta}$ the deviation from the ideal bond angle, $\frac{\boldsymbol{k_i}}{\boldsymbol{2}}$ is the associated force constant, and $\frac{\boldsymbol{V_n}}{\boldsymbol{2}}$, **ω**, and **γ** are the amplitudes, dihedral angle and phase angle associated with Fourier terms. Two forms are used for the four-body dihedral angle interactions: a harmonic term for improper dihedral angles **ζ**, that are not allowed to make transitions (i.e. dihedral angles in aromatic rings) and a sinusoidal term for the proper dihedral angles **ω**, which make 360 degree turns.[254] The non-bonded (intermolecular) interactions are modeled with a sum of over all pairs of atoms composed of van der Waals (Lennard-Jones) potential to account for dispersion and repulsion terms and Coulomb (electrostatic) potential to account for partially charged atoms:

$$E_{nonbonded} = E_{van\ der\ Waals} + E_{electrostatic}$$

$$= \Sigma_{i=1}^{N} \Sigma_{i=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right)$$

(Equation 2.22 *Non-bonded terms*)

where the van der Waals term expresses the interaction energy between two atoms (containing both repulsive r^{12} and attractive r^6 parts), r_{ij} represents the distance between atoms *i* and *j* with charges $\boldsymbol{q_i}$ and $\boldsymbol{q_j}$, $\boldsymbol{\varepsilon}$ represents the dielectric constant. The calculation of Coulomb interactions are the most time consuming as it is evaluated by means of a pairwise additive potential; therefore the number of contributions for N-particles of a system is N(N-1)/2 and scales as N^2.[254] As this represents the highest percentage of computational resources employed in MD simulations, many efficient algorithms have been developed to save computational time in the evaluation of this term.

2.8.3 Computational efficiency methods

2.8.3.1 PME method

Due to the prohibitive computational cost and difficulty it takes to calculate long-range electrostatic pairwise interactions several techniques have been developed. The most popular and efficient technique is known as the **Particle-Mesh-Ewald (PME) method**[255] which calculates long-range forces within a finite distance using a modification of Coulomb's Law and in reciprocal space using a Fourier transform to build a "mesh" of charges, interpolated onto a grid.[255] The fast Fourier algorithm scales as *N lnN*, which gives considerable advantages over the N^2 alternative.[250]

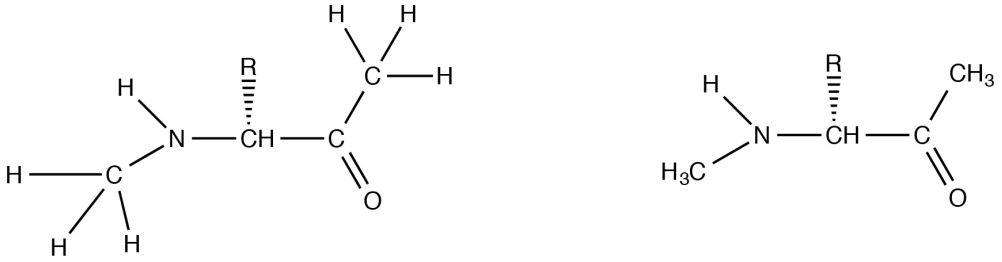
The Fourier coefficient of the mesh-based charge density is computed and then the potential on each particle is transformed back incorporating them into the non-bonded interactions. This method is capable of achieving high accuracy and is markedly efficient as it

evaluates the long-range forces by analytically differentiating the energies, thus reducing memory requirements substantially with relatively little increase in computational cost.[256]

2.8.3.2 United atom force field

As the number of non-bonded interactions scales with the squared number of interaction sites, the computational expense also greatly increases.[250] It is therefore advantageous to reduce the number of interaction sites in order to save computational time by employing a united atom model where several atoms are joined into a single interaction site. In this approach, non-polar hydrogens atoms in methyl (CH_3) and methylene (CH_2) groups do not have explicit representation and instead are represented by a single united atom.

Figure 2.5 United atom representation



Adapted from Leach, A. R., 2001[250]

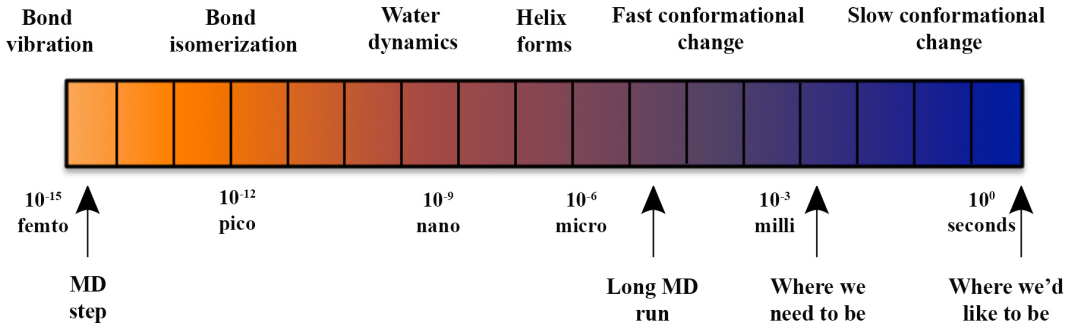
In this way, not only are the explicitly simulated degrees of freedom reduced, but also the number of atoms in the system. This yields a reduction of computational effort up to a factor of nine at the expense of neglecting the slight directional and volume effects of the presence of the hydrogens.[257] The forces between the interaction sites act on the united atom mass center, which remains located on the carbon atom (with a mass of 15 for a CH_3 group and 14 for a CH_2

group).[250] Polar hydrogens are represented explicitly so that they are available to participate in hydrogen-bonding interactions.

2.8.3.3 Constraints and restraints

Another way to increase simulation efficiency is to apply constraints or restraints to bond lengths or angles. Since MD simulations involve three time scales: the integration time step (δt), the total simulation time t, and the time scale of the process of interest τ_r, the time step needs t to be small enough so the changes in the forces over one step are small in order to describe the correct motion of the particles in a system.[258] Moreover, the length of the time step in a MD simulation is limited by high frequency internal motions occurring in the system (**Figure 2.6**).[254] In freezing the high vibration frequencies by constraining the bonds lengths or angles, vibrations are removed; therefore increasing the time step and computational efficiency is increased by a factor of 3.

Figure 2.6 Vibrational frequency timescale



Constraints or restraints have been developed for the constraining of the fastest degrees of freedom. In **constraint** dynamics, bonds or angles are forced to adopt specific values (fixed)

throughout a simulation without affecting the other internal degrees of freedom, whereas in applying **restraints**, the bonds or angles have the ability to deviate from the desired value.[250] When constraints are applied to all bond lengths or those involving only hydrogen atoms, stable trajectories with δt = 2 fs can be obtained.[258] When restraints are placed upon a bond or angle, it is able to deviate from the desired value; the restraint only acts to 'encourage' a particular value.[250] Typically both distance (length) and dihedral (angle) constraints are optional terms of the potential energy function (**see Equation 2.22**) and are introduced using harmonic interactions:

$$E_{Cr} = \Sigma K_i (r_i - r_{io})^2 \qquad \text{(Equation 2.23 } \textit{Constraint distance}\text{)}$$

$$E_{C\phi} = \Sigma K_i (\phi_i - \phi_{io})^2 \qquad \text{(Equation 2.24 } \textit{Constraint dihedral}\text{)}$$

From Becker, O.M. and Karplus, M.[259]

where $\boldsymbol{K_i}$ are the 'force' constants and $\boldsymbol{r_{io}}$, $\boldsymbol{\phi_{io}}$ are the target distance and dihedral values, respectively.

There are several methods available to apply constraints; the more common are the SHAKE[260] and LINear Constraint Solver (LINCS)[261] algorithms. In the **SHAKE**[260] algorithm, (iterative, slow) bond lengths or angles are constrained at their ideal values so that their stretching motions are not integrated prior to applying the rigid constraints.[259] The **LINCS** algorithm (non-iterative, fast) directly resets the constraints rather than the derivatives of the constraints.[262] A critical aspect to consider is that constraints are assessed as additional energy terms, acting as 'penalty functions'; if they are not satisfied, the energy increases significantly.[259] Therefore it is necessary to turn off the constraints as the energy of the conformation is calculated.

2.8.4 Water models

Molecular dynamics simulations involve a large number of water molecules and despite their small size, the calculation of non-bonded interactions between pairs of molecules and water accounts for most of the computational time. Most force fields utilize effective pairwise potentials to simulate liquid water, which then determines the computational efficiency.[250] Several water models have been developed and consist of three types: simple, flexible, and polarizable. **Simple** models represent a fixed (rigid) atom position where the interaction between the molecules is described using Coulombic and Lennard-Jones expressions. **Flexible** models represent atoms on a "spring" and enable internal changes in conformation. **Polarizable** models account for explicit polarization and many-body effects.

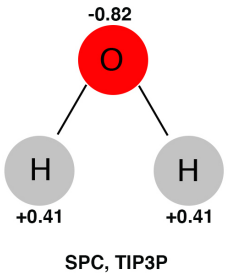
Simple water models are the most commonly used and consist between three to five interaction sites, which differentiate in the way that partial charges are distributed within the molecule.[263] Regardless of the number of sites, it is important that a single Lennard-Jones interaction site is located on the oxygen atom.[263] Common models with three interaction sites for the electrostatic interactions are SPC[264] and TIP3P[265] where the hydrogen atoms have partial positive charges that are balanced by a negative charge on the oxygen atom. The Lennard-Jones potential accounts for the van der Waals interaction between two water molecules based on the oxygen atom only and the Coulombic term accounts for the electrostatic interaction. The potential is given by the following equation:

$$= \Sigma_{i=1}^{N} \Sigma_{i=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) \qquad \text{(Equation 2.25 } \textit{Potential}\text{)}$$

where $\boldsymbol{\sigma_{ij}}$ is the collision diameter, $\boldsymbol{\varepsilon_{ij}}$ is the well-depth, $\boldsymbol{\varepsilon_0}$ is the dielectric constant, and $\boldsymbol{r_{ij}}$ is the distance between two charged atoms $\boldsymbol{i}$ and $\boldsymbol{j}$ with partial charges $\boldsymbol{q_i}$ and $\boldsymbol{q_j}$. The effective dipole moment of the SPC[264] model is 2.27 D and 2.34 D for TIP3P[265], which are close to the dipole

moment of liquid water (~2.6 D). Both models demonstrate good agreement between experiment and theoretical calculations of water with SPC[264] offering faster diffusion and TIP3P[265] faster dynamics.[263] The geometry and designated charges of the SPC[264] and TIP3P[265] water models are as follows:

Figure 2.7 SPC and TIP3P 3-site water model

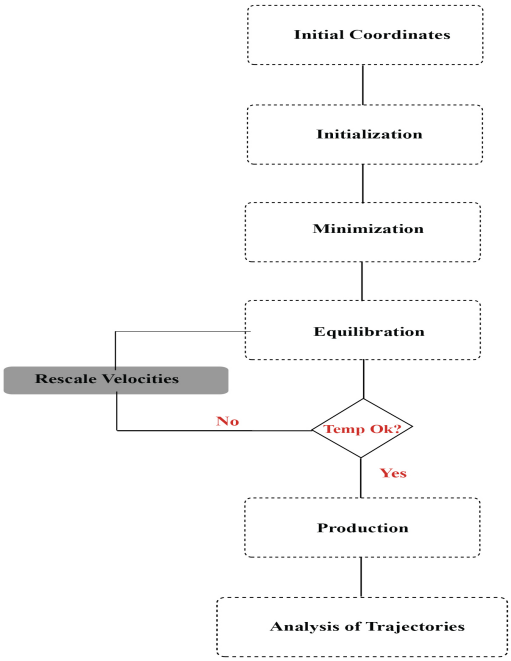


Adapted from Leach, A. R., 2001[250]

2.8.5 Setting up MD simulations

Molecular dynamics simulations involve numerically integrating equations of motion for a many-body system. When using Cartesian coordinates, one can integrate Newton's second law of motion to obtain the dynamical trajectory for each particle. Therefore, to begin a MD simulation an initial configuration (positions and velocities) of the system of interest must be chosen based on experimental or theoretical data such as a homology model, NMR structure, or X-ray crystal structure obtained from the Protein Data Bank (PDB).[205] The initial configuration must be chosen carefully as this can influence the quality of the simulation. The simulation itself can be broken down into several steps:

Figure 2.8 MD simulation steps



Choosing an appropriate time step, δt, must also be carefully considered; not too short so that conformations are sampled and not too long to prevent uncontrolled fluctuations. An order of magnitude less than the fastest motion is typically ideal - typically bond stretching is fastest motion, ~10 fs; therefore 1-2 fs is recommended. The **initialization** phase proceeds when initial velocities, $\boldsymbol{v_i}$, are randomly assigned to atoms to generate a Maxwell-Boltzmann distribution which gives the probability, $\boldsymbol{p(v_{ix})}$, that an atom $\boldsymbol{i}$ has a velocity, $\boldsymbol{v_x}$, in the $\boldsymbol{x}$ direction at a given temperature:

$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_B T}\right)^{1/2} exp\left[-\frac{1}{2}\frac{m_i v_{ix}^2}{k_B T}\right] \qquad \text{(Equation 2.26 } \textit{Maxwell-Boltzmann distribution}\text{)}$$

From Leach, A. R., 2001[250]

Once the initial velocities are assigned, **energy minimization** is subsequently performed which allows water molecules to adjust to the protein and removes any existing strong van der Waals interactions which might otherwise lead to local structural distortion and result in an

unstable simulation. Following energy minimization, the system is then **equilibrated** where the temperature is systematically increased from 0 K to a desired temperature (i.e. depends on system: 273 K for room temperature or between 300-310 K if body temperature is desired). During this phase, the initial velocities are assigned at a low temperature and gradually scaled upward by means of a thermostat; therefore allowing the process to continue until the desired temperature and a Maxwell-Boltzmann distribution is achieved. Additionally, the integration step is appointed where the force on each atom is calculated by differentiating the potential function. Thermodynamic and structural properties are monitored throughout equilibration until stability is achieved.

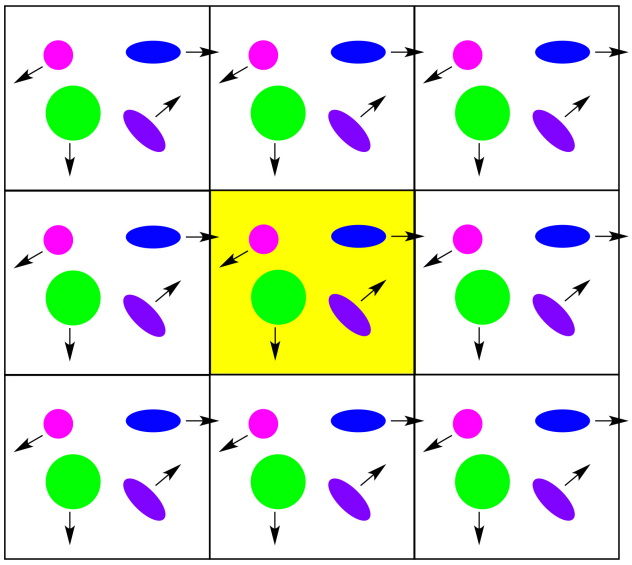
Once equilibration has been attained at the desired temperature and pressure, the final **production** phase is initiated and carried out in a desired thermodynamic ensemble (i.e. isobaric-isothermal, NPT ensemble, where the number of atoms of atoms, N, pressure, P and temperature, T are fixed). During this phase thermodynamic properties are derived from Newton's to obtain atomic positions and velocities as a function of time.

2.8.6 Periodic boundary conditions

Boundary effects play a critical role in MD simulations since the size of a system is finite; therefore the correct treatment of boundaries are necessary to minimize edge effects and avoid surface artifacts. Periodic boundary conditions (PBC) enable a simulation to be performed using a relative small number of particles in such a way that the particles experience forces as if they were in bulk fluid.[250] When applying PBC the atoms of the system are placed in a pre-defined box (cubic or rectangular), or more generally into a periodic space-filling box, which is surrounded by identical, translated copies of itself. When an atom exits the central box at one side, it enters it with identical velocity at the opposite side at the translate image position. This process is intimately related to non-bonded interactions, where the forces are summed over all

neighbors in the infinite periodic system and only coordinates from the central box are stored throughout the simulation.

Figure 2.9 Periodic boundary conditions



2.8.7 NPT ensemble

When performing MD simulations, it is critical to keep the certain state variables (i.e. temperature and pressure) of the system constant throughout the duration of the simulation in order to mimic previous experimental (macroscopic) conditions. Depending on which state variables are kept fixed (i.e. energy E, volume V, temperature T, Pressure P, and number of particles N), different statistical ensembles can be generated (**Table 2.3**). A variety of properties (structural, energetic, and dynamic) can then be calculated from the averages or fluctuations of these quantities over the ensembles.

The constant-pressure, constant-temperature (NPT) ensemble can be employed throughout the production phase of a MD simulation, as it compares with experimental results where conformational changes, protein-ligand and drug-target binding have been modeled under

constant temperature and pressure. The NPT ensemble is maintained by coupling the system to an external 'heat bath' (thermostat), at a fixed temperature.[266] Similarly to maintain pressure, the system is coupled to an external 'pressure bath' (barostat). The baths supply or remove heat or pressure from the system as needed and scales the velocities at each time step.

Table 2.3 Thermodynamic ensembles

Thermodynamic Ensembles Select one from each row	
Particle number, ***N***	Chemical potential, $\boldsymbol{\mu}$
Volume, ***V***	Pressure, ***P***
Energy, ***E***	Temperature, ***T***

2.8.8 Analysis

Molecular dynamics simulations can be analyzed by extracting the desired properties (i.e. atomic positions, velocities, and forces as a function of time) to gain insights into protein function, conformational dynamics, correlated motions, hydrogen-bonding, water-mediated interactions, and secondary structure changes. The most common analyses performed on MD simulations are discussed below, each of which yields a time series of values derived from the trajectory.

2.8.8.1 Root-mean-square deviation (RMSD)

The most commonly used quantitative measure of the similarity between two superimposed models is typically calculated by the root-mean-square deviation (RMSD):

$$RMSD = \sqrt{\frac{1}{N}\Sigma_{j=1}^{N}\left(ri_j - r0_j\right)^2} \qquad \text{(Equation 2.27 } \textit{RMSD}\text{)}$$

where $\boldsymbol{r_i}$ represents the position at time $\boldsymbol{i}$, $\boldsymbol{r_0}$ represents position at time 0, and $\boldsymbol{N}$ represents the number of atoms. RMSD quantifies the deviation of a protein from the reference conformation as

a function of time. An optimal overlay can be based on all atoms overlapping or just the backbone, favoring a minimal RMSD overall. The RMSD can also be used as an indicator of convergence of the structure towards an equilibrium state.

2.8.8.2 Root-mean-square fluctuation (RMSF)

The root-mean-square fluctuation (RMSF) is a measurement of the thermal motions of a residue that captures overall atomic fluctuations, offering insight into the flexible regions of a protein. It is calculated by:

$$RMSF = \sqrt{\frac{1}{N}\Sigma_{i=1}^{N}(r_i - < r >)^2} \qquad \text{(Equation 2.28 } RMSF\text{)}$$

where $\boldsymbol{r_i}$ represents the position at time $\boldsymbol{i}$ and is a selectable parameter that represents the atoms, backbone or any other subset, $\boldsymbol{N}$ represents the number of samples, and $< \boldsymbol{r} >$ represents the average value. Typically the reference position will be the time-averaged value of the same particle.

2.8.8.3 Radius of gyration (Rg)

The radius of gyration (Rg) calculation provides an indication of the shape (compactness) of a protein as a function of time. It is directly connected to the folding nature of a protein. If a protein is folded, it will maintain a relatively steady value of Rg; if it unfolds, its Rg will change over time. The Rg is calculated by:

$$Rg = \sqrt{\frac{1}{N}\sum_i (r_i - r_{cm})^2} \qquad \text{(Equation 2.29 } Rg\text{)}$$

where $r_i - r_{cm}$ is the distance between atom i and the center of mass, r_{cm} of the molecule. Rg quantifies the overall compactness of a protein.

2.8.8.4 Clustering

The data generated from MD simulations is quite extensive giving rise to a challenging analysis; therefore a method to identify conformational states, which are frequently and repeatedly populated, is necessary. Clustering analysis is a popular and widely utilized data-reduction method that focuses on separating ensembles into groups of structures that share similar conformational features.[267] Therefore the representative number of structures can then be utilized for subsequent analysis rather than the thousands of conformations generated in a typical simulation. The goal of clustering is to assign patterns to a particular group. The similarity measure between two conformations used in clustering is based on the Euclidean distance and is calculated by:

$$d_{ij} = \sqrt{\sum_{m=1}^{N_{tor}} (\omega_{m,i} - \omega_{m,j})} \qquad \text{(Equation 2.30 } \textit{Clustering}\text{)}$$

where $\omega_{m,i}$ is the value of torsion angle m in conformation i and N_{tor} is the total number of torsion angles. Clustering can lower the complexity of the structural information contained in the many conformations. When comparing conformations, the RMSD is calculated and a similarity matrix is built. There are a large number of cluster algorithms that belong to the following categories:

1) hierarchal (top-down)
2) linkage (bottom-up)
3) Jarvis-Patrick
4) refinement

The three most commonly used cluster algorithms are the GROMOS,[268] linkage and Jarvis-Patrick[269] methods. The **GROMOS method**[268] utilizes the *hierarchal (top-down)* approach, where the neighbors of a structure are determined based on a given RMSD cut-off. The structure with the greatest number of neighbors is utilized as the center of the cluster and subsequently eliminated from the pool of structures. The process is repeated until the pool is empty. Clusters are mutually exclusive; therefore a structure can only be a member of a single cluster. The **linkage method** utilizes the *bottom-up* approach, where at the start of the analysis the data set contains as many clusters as conformations. Initially, all the inter-cluster distances are calculated between each pair of conformations. In each subsequent step, the closest pair of clusters merges into one until the closest pair differs by a predefined RMSD cut-off. The **Jarvis-Patrick method**[269] measures the distance between two objects and two integers, *J* and *K*, where *J* is the size of the neighbor list and *K* is the number of common neighbors. The nearest neighbors are determined for each object in the set to be clustered. Two conformations are assigned to the same cluster if: (1) the two conformations are within the cut-off distance and (2) they have the same number of nearest neighbors in common.

2.8.8.5 Principal component analysis (PCA)

Another data-reduction and frequently utilized technique is principal component analysis (PCA).[270-273] Conformational trajectories reside in high dimensional space containing information about all atomic coordinates. PCA is commonly used to reduce the dimensionality of a data set.[250] It can be used to calculate the directions and amplitudes of greatest motion along a simulation trajectory. In general, a principal component is a linear combination of the variables:

$$p_i = \sum_{j=1}^{v} c_{i,j} x_j \qquad \text{(Equation 2.31 } PCA\text{)}$$

with $\boldsymbol{v}$ variables, where $\boldsymbol{p_i}$ is the ith principal component and $\boldsymbol{c_{i,j}}$ is the coefficient of the variable $\boldsymbol{x_j}$. PCA is an orthogonal linear transformation that transforms a multi-dimensional data set into a reduced space spanned by principal components (PCs)[273]; thus highlighting principal motions or dominant features. It maps data sets of high dimensionality, for example, the vast vector space (**V**) spanned by the *3N* coordinates of a simulation system, sampled across millions of times steps (the matrix, **R**) into a new vector set (**V'**), defined by an alternative basis set. The application of PCA involves computing the variance of the different variables in a data set and subsequently arranging the data in descending order. The first principal component (PC1) contains the most variance; the second (PC2) contains the second highest variance and so forth. PCA begins by removing the six rigid-body translational and rotational degrees of freedom of the molecule via least-squares structural superimposition of each frame to a reference (i.e. initial structure). The following steps express the transformation of a *3N x m* matrix, **R** from *3N* Cartesian coordinates at frames 1,2,…,*m*. PCA is then achieved in two steps: (1) by using **R** to construct the variance-covariance matrix, **C**, of 3D coordinate displacements, $\vec{r}$ versus the trajectory-averaged mean coordinates $\langle\vec{r}\rangle$, and then (2) diagonalizing **C** to obtain the principal components of the motion as projections onto the eigenvector of the covariance matrix.

A data set is introduced and centralized by computing the mean:

$$S = \{x_1, \ldots, x_m\}$$

$$\bar{x} = \frac{1}{m}\sum_{i=1}^{m} \tilde{x}_i$$

(Equation 2.32 *PCA Mean*)

Construction of the variance-covariance matrix, **C**:

(1) $$C = RR^T = (\vec{r}(t) - \langle\vec{r}\rangle)(\vec{r}(t) - \langle\vec{r}\rangle)^T$$

(Equation 2.33 *PCA variance covariance matrix*)

The covariance matrix diagonalized:

(2) $$C = Y\Lambda Y^T$$

(Equation 2.34 *PCA matrix diagonalized*)

where **Y** is the orthogonal transformation of diagonalized matrix, **Λ** is diagonal matrix containing the corresponding eigenvalues (λ), and **T** denotes the transpose.

The solved eigenvalues are represented as:

$$Cv_i = \lambda_i v_i$$ (Equation 2.35 *PCA eigenvalues*)

and placed in decreasing order from λ_1 to λ_N . The eigenvectors correspond to the computed eigenvalues and are represented as:

$$V = (v_1, v_2, \dots, v_N) \qquad \text{(Equation 2.36 } \textit{PCA eigenvectors}\text{)}$$

Therefore, the original trajectory coordinates in the matrix R are transformed to a new coordinate system, which can be projected onto the eigenvectors (modes) in order to visualize the motion along teach of those directions giving the corresponding principal components. The use of PCA has been shown to be a reliable method to analyze protein-structure ensembles.[274,275]

Chapter 3: Structural mutation analysis of PTEN and its genotype-phenotype correlations in endometriosis and cancer

3.1 Introduction

The phosphatase and tensin homolog deleted on chromosome ten (*PTEN*) gene encodes a tumor-suppressor phosphatase that has recently been found to be frequently mutated in patients with endometriosis, endometrial cancer, and ovarian cancer. Here, we present the first computational analysis of 13 somatic missense *PTEN* mutations associated with these phenotypes. We found that a majority of the mutations are associated in conserved positions within the active-site and are clustered within the signature motif, which contain residues that play a crucial role in loop conformation and are essential for catalysis. *In silico* analyses were utilized to identify the putative effects of these mutations. In addition, coarse-grained models of both wild-type (WT) PTEN and mutants were constructed using elastic network models to explore the interplay of the structural and global dynamic effects that the mutations have on the relationship between genotype and phenotype. The effects of the mutations reveal that the local structure and interactions affect polarity, protein structure stability, electrostatic surface potential, and global dynamics of the protein. Our results offer new insight into the role in which *PTEN* missense mutations contribute to the molecular mechanism and genotypic-phenotypic correlation of endometriosis, endometrial cancer, and ovarian cancer.

The crystal structure of PTEN reveals two major domains: a lipid phosphatase domain that contains the catalytic active-site and a tightly associated C2 domain, which participates in membrane binding.[4] The wall of the active-site pocket is delimited in part by the signature motif

P loop ($H^{123}CKAGKGR^{130}$), WPD loop (residues 88-98), and TI loop (residues 160-171) which contain residues that are responsible for catalysis (D92, C124, and R130), the overall positive charge within the active-site (H93, K125, K128), mediation of loop motion (H123 and G127), and govern the depth and width of pocket (4-residue insertion: K163, K164, G165, and V166).[4] The tumor suppressor function of PTEN is dependent on its phospholipid phosphatase activity and the loss-of-function of the phosphatase catalytic domain is commonly associated with oncogenic *PTEN* mutations.[4,178-180] Recently, somatic mutations and deletions of *PTEN* have been reported in many types of sporadic tumors, including endometriosis, endometrial, and ovarian cancers.[147,156,176,182-189,276]

Despite over three decades of research and significant understanding of endometriosis, the molecular mechanism and pathogenesis underlying its proliferation remains incompletely understood. The incidence of *PTEN* mutations in endometrium tissue of women diagnosed with endometriosis and endometrial hyperplasia is one of the highest among analyzed tumors and the most commonly mutated gene identified in endometrial cancer.[169,182] Recent evidence suggests that endometrial and ovarian carcinomas are frequently found in association with endometriosis, suggesting that they arise through malignant transformation.[50,126,147,156,187-189,276] Moreover, women with a history of endometriosis are also at an increased risk for developing cancer.[126-129] These data suggest that a greater understanding of the complicated mechanism of endometriosis and malignant transformation will not only illuminate its molecular etiology, but may lead to both therapeutic development and early diagnostic methods.

Insights into the catalytic properties of PTEN provide a platform to analyze the mutations found in endometriosis, endometrial cancer, and ovarian cancer. Recent computational studies on PTEN reveal that missense mutations affect its function and structure.[192,193] Therefore a careful dissection of the effects of putative mutations is pertinent to understanding the molecular mechanism in each of the identified phenotypes. We analyzed thirteen somatic missense PTEN

mutations that were identified from published literature, and further validated them in the My Cancer Genome database[277] as well as the Catalog of Somatic Mutations in Cancer (COSMIC) database.[278] Our approach focuses on understanding the impact that these missense mutations have on the structure of PTEN and on investigating the molecular mechanism of each of the phenotypes.

3.2 Computational methods

3.2.1 PTEN somatic missense mutations dataset

Somatic missense mutations for endometriosis, endometrial cancer, and ovarian cancer were taken from both the My Cancer Genome[277] and the Catalogue of Somatic Mutations in Cancer (COSMIC) (version 49) databases[278] and a comprehensive literature screening was conducted as listed in **Table 3.1**. A cut-off of 13 distinct missense mutations were chosen based on the highest percentage and frequency of the individual mutation somatic mutations found in the endometrium and ovaries as indicated in **Table 3.2** and **Figure 3.1**.

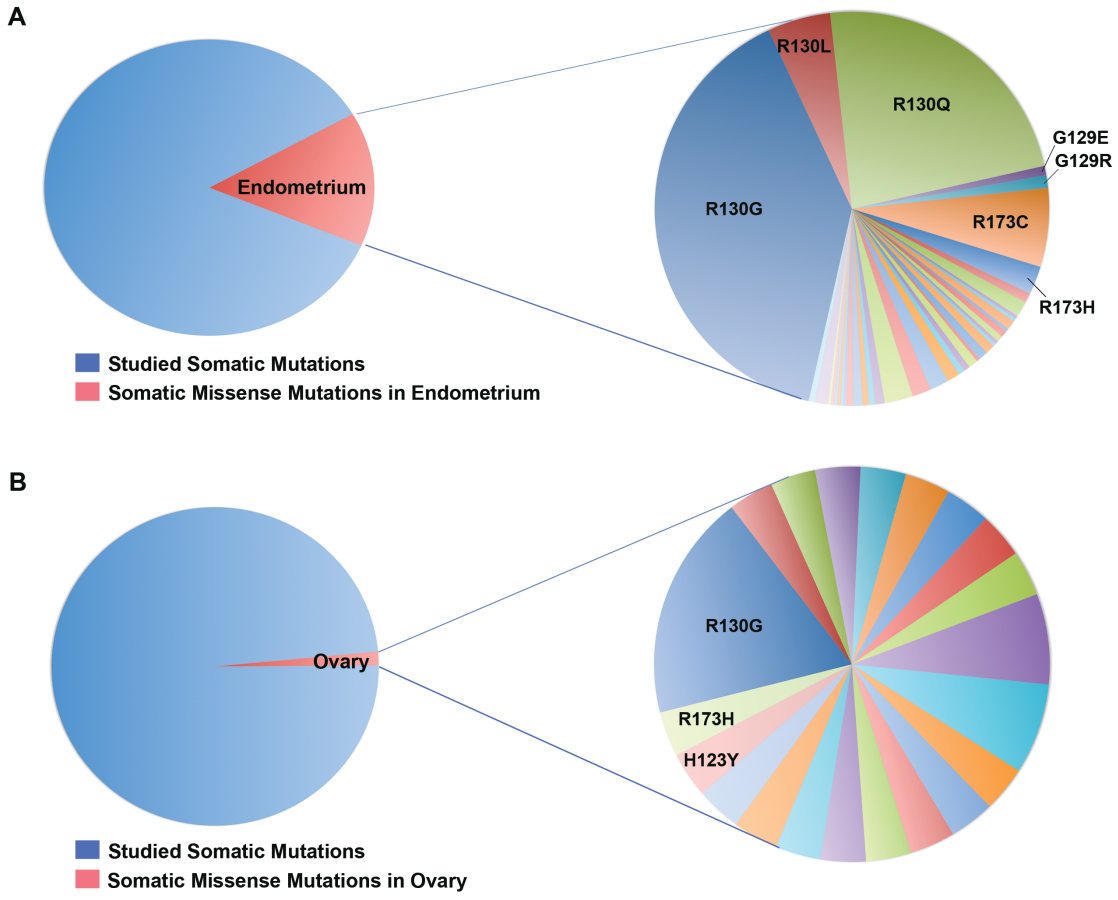
Table 3.1 PTEN somatic missense mutations and phenotypes

Mutations	Phenotype	Domain	Reference
p.G36R	Endometrial Hyperplasia	Phosphatase	Maxwell *et al.,* 1998; Minaguchi *et al.,* 2001
p.G36E	Endometrial Cancer		Dutt *et al.,* 2008
p.H123Y	Endometrial Cancer		Myers *et al.,* 1997; Bonneau *et al.,* 2000
p.C124S	Endometrial Cancer		Kurose *et al.,* 1998
p.G129R	Endometrial Cancer		Risinger *et al.,* 1997; Furnari *et al*., 1997
p.G129E	Endometrial Cancer		Risinger *et al.,* 1997; Weng *et al.*, 2001
p.R130G	Endometrial Cancer		Bonneau *et al.,* 2000; Obata *et al.,* 1998; Byron *et al.,* 2008; Dutt *et al.,* 2008; Oda *et al.*, 2005; Minaguchi *et al.,* 2001
	Endometriosis		Willner *et al.*, 2007
	Ovarian Cancer		Lovly *et al.*, 2015
p.R130Q	Ovarian Cancer		Sampson 1925; Sainz de la Cuesta *et al.,* 1996; Saito *et al.,* 2000; Castiblanco *et al.*, 2006; Mandai *et al.,* 2009; Xu *et al.,* 2011 Lovly *et al.*, 2015
p.R130L	Endometriosis		Maxwell *et al.,* 1998; Han *et al.,* 2000; Obata *et al.,* 1998
p.R173C	Endometrial Hyperplasia Endometrial Cancer		Maxwell *et al.,* 1998; Risinger *et al.,* 1998 Bussaglia *et al.,* 2000; Minaguchi *et al.,* 2001; Dutt *et al.*, 2008; Risinger *et al*., 1998
p.R173H	Ovarian Cancer		Sampson 1925; Sainz de la Cuesta *et al.,* 1996; Saito *et al.,* 2000; Castiblanco *et al.,* 2006; Mandai *et al.,* 2009; Xu *et al.,* 2011
p.V191A	Endometrial Hyperplasia	C2	Maxwell *et al.,* 1998; Han *et al.,* 2000
p.T348I	Endometrial Hyperplasia		Maxwell *et al.,* 1998; Georgescu *et al.,* 1999; Han *et al.,* 2000

Table 3.2 Incidence of PTEN mutations in cancers

Incidence of somatic missense *PTEN* mutations in human cancers	Percentage of tumors with mutation/number of samples	
	PTEN	
Primary Tumor Tissue Survey	**%**	**#**
Prostate	1	658
Breast	1	1198
Endometrium	9	1837
Cervix	0	413
Vulva	6	18
Soft Tissue	1	298
Upper aerodigestive tract	1	777
Biliary Tract	4	56
Ovary	1	845
Stomach	2	542
Liver	1	380
Kidney	1	474
Pancreas	2	198
Central Nervous System	4	3638
Large Intestine	1	1114
Meninges	1	170
Eye	6	71
Skin	1	768
Thyroid	0	715

Figure 3.1 Distribution of *PTEN* somatic mutations in endometrium and ovaries



3.2.2 Generation of PTEN structure mutants

The wild-type (WT) structure of human PTEN protein crystallized with tartrate (TLA) (PDB ID 1D5R) was obtained from the Protein Data Bank (PDB).[205] The tartrate molecule was removed and the calculations were conducted on apo PTEN. The mutant structures were built *in silico* by side chain replacement utilizing the Visual Molecular Dynamics (VMD) Mutator Plugin 1.3[207] starting from the WT PTEN (PDB ID: 1D5R) crystal structure file. Thirteen mutant structure models were generated for each of the phenotype mutations (p.G36E, p.G36R, p.H123Y, p.C124S, p.G129E, p.G129R, p.R130G, p.R130L, p.R130Q, p.R173C, p.R173H,

p.V191A, and p.T348I).

Energy minimization was performed on each PTEN mutant structure using GROMACS 4.6.3[270] employing the GROMOS53a6 force field[257] to remove possible steric clashes and minimize forces introduced as a result of the mutation. A series of five overall minimization stages were performed to remove any steric clashes and minimize the forces that were introduced as a result of the mutation that was introduced to the WT structure. Each PTEN mutant was subjected to a step-wise energy minimization using the steepest descent method. Step-wise energy minimization involved 5 steps: in step 1 all protein atoms were restrained, in step 2 protein heavy atoms were restrained, in step 3 all protein atoms were restrained again, in step 4 atoms of the protein main chain were restrained and in step 5, unrestrained energy minimization. Total minimization was carried out until convergence where the maximum atomic force was less than 1000 kJ/mol-nm.

3.2.3 pKa prediction and electrostatic surface potential (ESP) calculations

The single-site pKa predictions[199,279-281] were carried out on WT PTEN (Apo) at pH 7 using Chemistry at HARvard Molecular Mechanics (CHARMM),[282] and University of Houston Brownian Dynamics (UHBD).[202] The PTEN PDB entry 1D5R[4], contains the atomic coordinates of the protein as required for calculation of pKa. The pKa input contained partial atomic charges and atomic radii for each atom, maximal number of iterations [300], ionic strength [150 mM], dielectric constant [solvent 80, protein 20], temperature [298K], and spacing [4: 3.0 45 45 45; 1.2 15 15 15; 0.75 15 15 15; 0.25 20 20 20] in Angstroms.

Upon completion of the pKa calculation of each ionizable residue, an electrostatic surface potential (ESP) map was generated. Electrostatic interactions are long ranged interactions and

can accelerate the rate at which molecules associate.[208] The ESP calculations were carried out for both WT PTEN and mutants using Adaptive Poisson-Boltzmann Solver (APBS).[283]

3.2.4 Multiple sequence alignment (MSA) – sequence conservation

The protein sequence of PTEN was searched against the UniProt[284,285] sequence database to obtain homologous PTEN proteins sequences. Multiple sequence alignments (MSA) were constructed utilizing CLUSTALW[210] and manually inspected using Jalview.[211]

3.2.5 Structural stability prediction

I-Mutant 3.0 (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) is a support vector machine (SVM)-based tool that calculates protein stability changes upon mutation either from sequence or structural information.[240] We utilized the structure-based version of I-Mutant 3.0, which classifies the prediction into one of three classes: neutral mutation ($-0.5 \leq \Delta\Delta G \geq 0.5$ kcal/mol), large decrease (<-0.5 kcal/mol), or large increase (>0.5 kcal/mol).[240] The predicted free energy change, $\Delta\Delta G$, is based on the difference between unfolding Gibbs free energy change of mutant and native protein (kcal/mol).[240] **CUPSAT** (http://cupsat.tu-bs.de) is a physiochemical modeling program that utilizes a Boltzmann mean-force potential and torsion angles (phi, psi) to predict $\Delta\Delta G$ values and stability changes.[241] CUPSAT also calculates the overall stability change as well as the adaptation (favorable or unfavorable) of the observed torsion angles (phi, psi) as a result of the mutation. The WT *PTEN* structure file (PDB ID 1D5R)[4] was utilized with the residue name and mutation site required as input.

To explore the effects of the mutations on structure stability and dynamics we employed **Elastic network contact model (ENCoM)** (http://bcb.med.usherbrooke.ca/encom),[242,243] that utilizes a novel mixed coarse-grained normal mode method to account for the type and extent of

pairwise atomic interactions allowing for the calculation of vibrational entropy differences as a result of mutations.[242,243] ENCoM utilizes a potential function with four terms: (1) covalent bond stretching, (2) angle bending, (3) dihedral torsion, and (4) non-bonded interaction while taking into consideration the nature and possible effects that the orientation of the side-chains have on dynamics within the context of normal mode analysis. The WT PTEN structure file (PDB ID 1D5R)[4] was utilized with the residue name and mutation site required as input.

3.2.6 Anisotropic network normal mode analysis (NMA)

The anisotropic network model (ANM), is a coarse-grained model utilized to investigate protein dynamics and vibrational functional motions.[225] Utilizing elastic network methodology (ENM), ANM calculations involved only Cα atoms and employed a uniform spring constant γ with a cut-off distance of 1.5 nm for interacting atoms, such that the overall potential of the system is a sum of harmonic potentials.[223,286] Anisotropic displacements and intrinsic collective motions of WT and mutant PTEN structures were analyzed using ProDy[287] to obtain the six slowest nontrivial normal modes.

3.2.7 All-atom normal mode analysis (NMA)

All-atom normal mode analysis (NMA) was conducted on representative molecular dynamics (MD) simulations using GROMCAS 4.6.3[270] with GROMOS53a6 force field[257] on (apo) WT PTEN (PDB ID 1D5R) and mutants as a structural basis. Each system was solvated inside a cubic box of simple point charge (SPC)[264] water with at least 10 Å of water between the protein and edges of the box. All simulations were performed in explicit solvent, with chloride (Cl-) counter ions added to obtain neutrality of the system. Periodic Boundary Conditions and a 2 fs time step were employed for each simulation. Each system contained roughly 75,825 atoms.

The Particle Mesh Ewald (PME) method[288] was used to treat long-range electrostatic interactions and a cutoff of 9 Å was used for non-bonded interactions. A series of five overall minimization stages were performed on each PTEN model to remove any steric clashes and minimize the forces that were introduced as a result of the mutation introduced to the WT structure. Each system was then subjected to a step-wise energy minimization using the steepest descent method. The minimized structures were then slowly heated from 0 to 300 K over 100 ps and equilibrated for an additional 250 ps. The system was heated to 300 K by linearly by increasing the temperature, through velocity rescaling, every 10 ps. The production runs were performed in the NPT (isobaric-isothermal) ensemble at 300 K. Bond lengths were constrained using LINear constraint solver (LINCS) algorithm[261] and the van der Waals forces were maintained at 1.4 nm. The Berendsen[266] weak coupling method was employed to maintain constant temperature with a temperature coupling relaxation time of 0.1 ps, a pressure coupling constant of 0.5 ps, and a compressibility of $4.5 \bullet 10^{-5}$. The total simulation time for each model was 200 ns and coordinates were recorded every 1 ps. Normal mode analysis of each system was conducted by performing principal-component analysis. Normalizing each eigenvalue of the covariance matrix to its total sum yields the percentage of all movements attributable to the corresponding eigenvector. The largest eigenvalues correspond to the principal-component modes that best explain the molecular motions sampled by the all-atom system trajectories.

3.3 Results and discussion

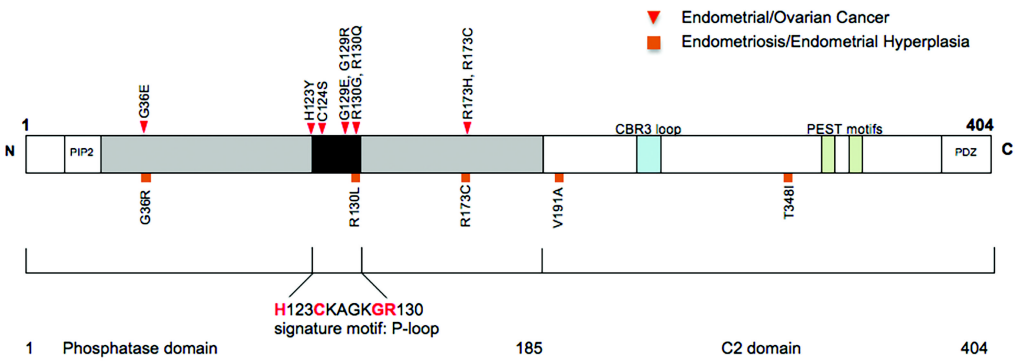
3.3.1 Mapping somatic missense mutations to PTEN structure

To investigate the pattern of *PTEN* mutations in each of the phenotypes, we analyzed the prevalence of *PTEN* somatic missense mutations from data presented in literature over the past three decades. The spectrum of their associated phenotypes is shown in **Table 3.1**. Using the My

Cancer Genome and the Sanger COSMIC databases, *PTEN* was screened and the thirteen mutations were confirmed to be within the highest percentage of somatic mutations found in the endometrium and ovaries (**Table 3.2** and **Figure 3.1**). Using these data, we investigated the functional impact of the mutations and whether any correlation exists between *PTEN* genotype and disease phenotype.

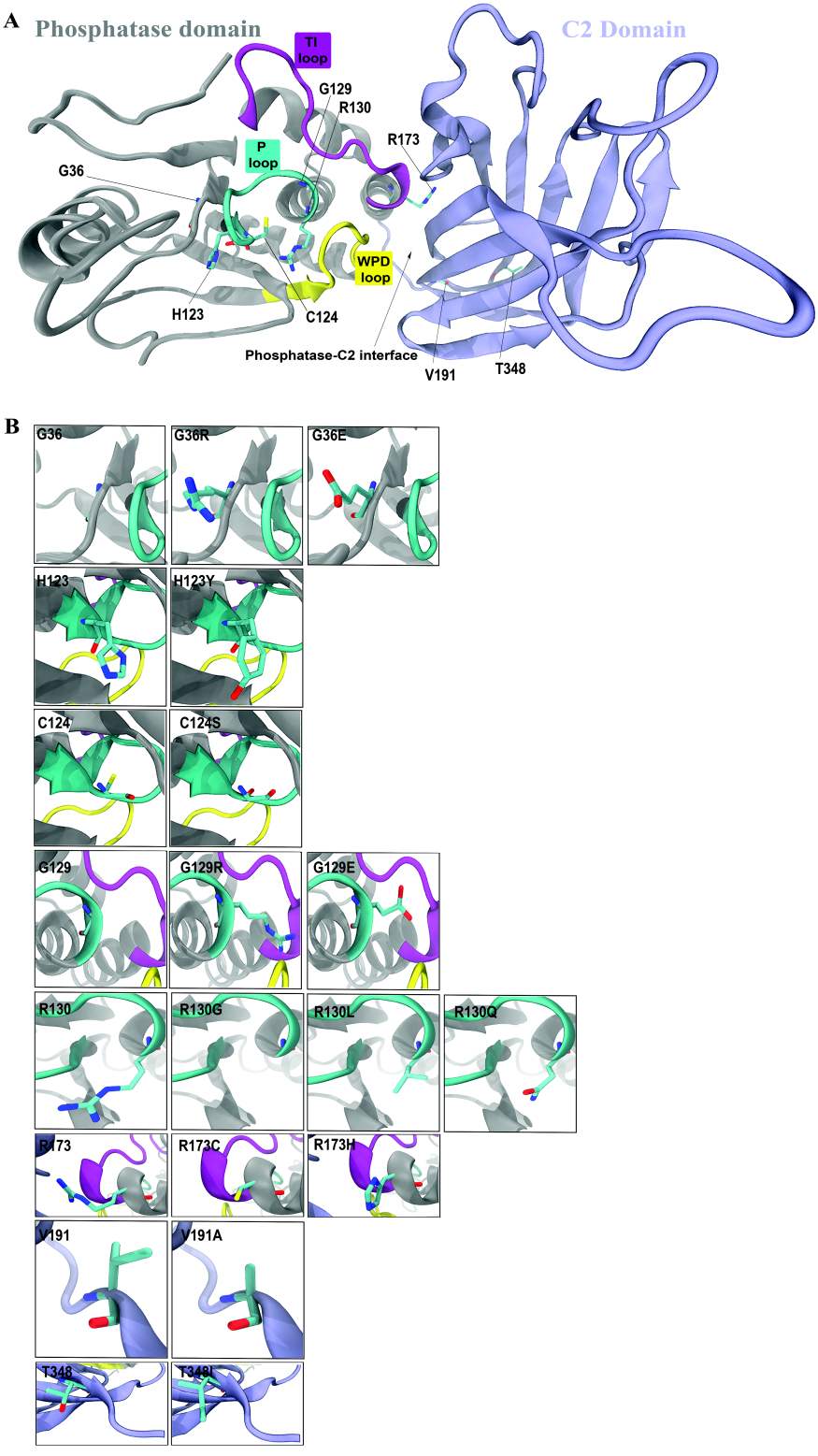
The domain structure of the PTEN protein comprises a phosphatase domain (residues 1-185) and a C2 domain (residues 186-351), which are both vital for tumor suppressor function. The phosphatase domain contains a tyrosine phosphatase signature motif ($H^{123}CKAGKGR^{130}$) that forms a P loop within the active-site pocket. Within this loop, the C124 and R130 residues are essential for catalysis. Knowledge of the catalytic mechanism provides a rational basis for understanding a number of the mutations reported within *PTEN*. Thus, we created a histogram, identifying mutational hotspots within the domain structures of PTEN to further highlight mutations surrounding the active-site that may affect catalytic activity and structural stability (**Figure 3.2**). We found the vast majority of mutations localized within the phosphatase domain, particularly the signature motif, suggesting that phosphatase activity is essential for the physiological function of PTEN in tumor suppression.

Figure 3.2 Mutation distribution within the domain structure of human PTEN



To further examine the local structure and interactions around each mutated residue in both domains of PTEN, these mutations were mapped to the three-dimensional structure of PTEN as shown in **Figures 3.3 A-B**. We observed that nine mutations associated with severe phenotypes: six (p.H123Y, p.C124S, p.G129E, p.G129R, p.R130G, and p.R130Q) are clustered within the active-site pocket, one (p.G36E) is located near the N-terminus and two (p.R173C and p.R173H) are located at the phosphatase-C2 domain interface. Four of the thirteen mutations (p.G36R, p.R130L, p.V191A, and p.T347I) present mild phenotypes and are scattered throughout both PTEN domains (**Figures 3.2** and **3.3 A-B**). p.R173C is involved in both mild and severe phenotypes: endometrial hyperplasia, endometrial cancer, and ovarian cancers, respectively (**Table 3.1**).

Figure 3.3 Three-dimensional structure of PTEN and mutants



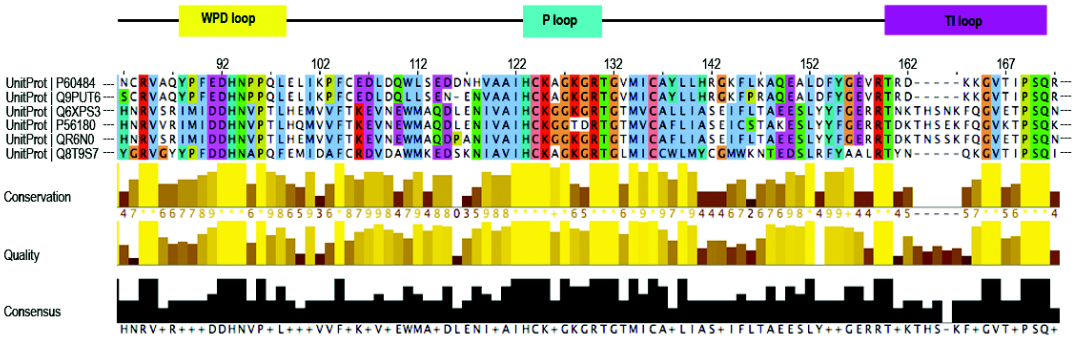
Residues C124 and R130 are affected by missense mutations and play significant roles in catalytic activity and protein function. Most importantly, R130 has mutations (R130G, R130L, and R130Q) implicated in each of the three phenotypes: endometriosis, endometrial cancer, and ovarian cancer (**Table 3.1** and **Figures 3.3A-B**). These three distinct R130 mutations considerably change the size of the respective amino side chains such that the local structure and interactions might be disturbed. Three loops (WPD, P, and TI) that comprise the active-site pocket of the phosphatase domain are responsible for the overall positive charge (H93, K125, and K128), catalysis (D92, C124, and R130), mediation of loop motion (H123 and G127) and govern the depth and width of the pocket (K163, K164, G165, and V166) (**Figure 3.3A**).[4]

Interestingly, the p.G129E mutation is found in patients with endometrioid tumors with high frequency.[30,179] The position of this residue is located at the bottom of the active-site cleft near the TI loop that contributes to the extended active-site pocket (**Figures 3.3A-B**). Mutations p.G129E and p.G129R reduce the size of the catalytic pocket, thus hindering accommodation of the bulky phosphoinositol moiety of the substrate PIP_3 (**Figures 3.3A-B**).[4] This mutant likely retains protein tyrosine phosphatase (PTP) activity because the phosphorylated inositol head group requires a larger and shallower pocket.[289] Additionally, the glycine is located in a conserved region and is essential at this position to adopt a backbone conformation needed to accommodate the phosphoinositol moiety and a mutation to glutamic acid thereby reduces the flexibility of the backbone (**Figure 3.4**). Substitution of arginine (p.G129R), which has a larger side chain than glutamic acid, results in complete loss of both protein tyrosine phosphatase and lipid phosphatase activities.[289] It is possible that mutations involving the P, WPD, or TI loops disrupt the size and overall positive charge of the active-site pocket, the tumor suppressor function of PTEN, and thus contributes to the development of tumors with a more severe phenotype.[290]

3.3.2 PTEN sequence conservation

Disease-causing mutations are typically located at conserved positions within a protein family since these residues are usually essential for the structure and/or function of the protein.[291,292] Within PTEN, the P loop is highly conserved between species, whereas the WPD and TI loops are more divergent (**Figure 3.4**). The inter-domain interface residues of the phosphatase domain are second only to the P loop in conservation across species and those of the C2 domain are the best-conserved C2 regions (**Figure 3.3A**).[4] The majority of the somatic missense mutations investigated in this study are located within a contiguous conserved cluster (**Figure 3.2**) formed around the active-site and highly conserved P loop as well as the inter-domain region conferring significant phenotypic changes and can be targeted as potential hot-spot regions for the development of novel therapeutic agents in the treatment of endometriosis, endometrial cancer, and ovarian cancer.

Figure 3.4 Multiple sequence alignment (MSA) of PTEN catalytic loops

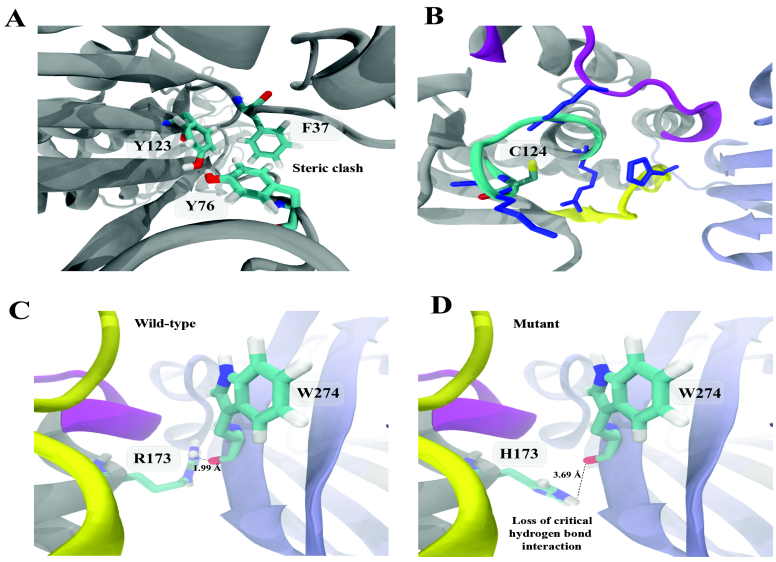


3.3.3 Structural and stability effects of PTEN mutations

To understand the structural and phenotypic consequences of the missense mutations in PTEN protein, we utilized a combination of molecular modeling, structural analysis, and bioinformatics methods to analyze the identified mutations. The effects of the mutations reveal

that the local structure and interactions affect polarity, protein structure flexibility, and electrostatic surface potential. Structural analysis of p.H123Y, p.C124S, and p.R173H reveal that alterations in the residue size and hydrophobicity disturb the structure and disrupt hydrogen bonding (**Figures 3.5A-D**). The signature motif contained in the loop between pβ5-strand and α4-helix includes two residues H123 and C124 essential for the conformation of the P loop and for catalysis, respectively. The mutation of H123 to tyrosine, causes a steric clash with Y76 disrupting the π-π stacking between H123 with Y76 and F37 thus extending the pβ5-strand before the P loop (**Figure 3.5A**).

Figure 3.5 Structure analysis of WT PTEN



Basic residues H93, K125, K128, surround the catalytic C124 residue and R130, and hydrogen bonds with the side chains R130 and T131 (**Figure 3.5B**). The p.C124S mutation introduces a slightly less hydrophobic side chain disrupting the hydrogen-bonding interactions, lowering the pKa (**Table 3.4**) and presumably its ability to act as a nucleophile with its known substrate $PI(3,4,5)P_3$. In fact, previous research has shown that p.C124S completely inhibits

PTEN's lipid and protein phosphatase activity and was defective in phosphatase activity against $PI(3,4,5)P_3$ and $PI(3,4)P_2$.[293,294] The p.R173H mutation disrupts inter-domain hydrogen bonding with W274 (2.2Å to 3.6Å) and imposes steric clash and side chain distortion with the nearby residue, V262 (**Figures 3.5C, D**).

Individual missense mutations may destabilize a protein and impact its folding ability. To consider this effect and estimate the level of structural destabilization associated with each missense mutation, we predicted quantitative stability changes utilizing multiple programs that calculate the change of the protein structural stability induced by mutations (ΔΔG). In order to provide points of comparison, we utilized multiple programs such as I-Mutant 3.0,[240] Cologne University Proteins Stability Analysis Tool (CUPSAT),[241] and Elastic Network Contact Model (ENCoM) (**Table 3.3**).[242,243] I-Mutant 3.0 was utilized to calculate protein structure stability changes from the predicted free energy change (ΔΔG) upon mutation as provided from either sequence or structural information.[240] I-Mutant 3.0 predicts each of the thirteen mutations (p.G36R, p.G36E, p.H123Y, p.C124S, p.G129R, p.G129E, p.R130G, p.R130L, p.R130Q, p.R173C, p.R173H, p.V191A, and p.T348I) to be destabilizing presenting with negative free energy change (ΔΔG) values -1.30, -0.94, -0.67, -1.12, -1.33, -1.03, -2.15, -0.56, -1.56, -1.57, -1.71, -0.18, and -1.61 kcal/mol, respectively. Eight of these mutations (p.G36R, p.H123Y, p.G129E, p.G129R, p.R130G, p.R173C, p.V191A, and p.T348I) are in agreement with experimental data that suggests the loss of lipid phosphatase activity toward PIP_3 as a result of the loss of stability in the protein.

Table 3.3 Predicted effects on protein stability (ΔΔG) for the 13 *PTEN* missense mutations

	I-Mutant 3.0		CUPSAT (thermal)			ENCoM	
Mutation	ΔΔG kcal/mol	Predicted Impact	ΔΔG kcal/mol	Predicted Impact	Torsion Angle	ΔΔG kcal/mol	Predicted Impact
p.G36E	-1.30	Destabilizing	+2.04	Stabilizing	Unfavorable	-0.54	Stabilizing
p.G36R	-0.94	Destabilizing	-2.77	Destabilizing	Unfavorable	-0.91	Stabilizing
p.H123Y	-0.67	Destabilizing	-2.13	Destabilizing	Favorable	-0.44	Stabilizing
p.C124S	-1.12	Destabilizing	+3.09	Stabilizing	Unfavorable	-0.17	Stabilizing
p.G129E	-1.33	Destabilizing	-2.30	Destabilizing	Favorable	-0.73	Stabilizing
p.G129R	-1.03	Destabilizing	-2.83	Destabilizing	Favorable	-1.15	Stabilizing
p.R130G	-2.15	Destabilizing	-2.21	Destabilizing	Unfavorable	+1.45	Destabilizing
p.R130L	-0.56	Destabilizing	+0.66	Stabilizing	Favorable	+0.23	Destabilizing
p.R130Q	-1.56	Destabilizing	+0.23	Stabilizing	Favorable	+0.27	Destabilizing
p.R173C	-1.57	Destabilizing	-0.27	Destabilizing	Unfavorable	+0.53	Destabilizing
p.R173H	-1.71	Destabilizing	+0.65	Stabilizing	Unfavorable	+0.12	Destabilizing
p.V191A	-0.18	Destabilizing	-1.39	Destabilizing	Unfavorable	+0.32	Destabilizing
p.T348I	-1.61	Destabilizing	-0.53	Destabilizing	Unfavorable	-0.86	Stabilizing

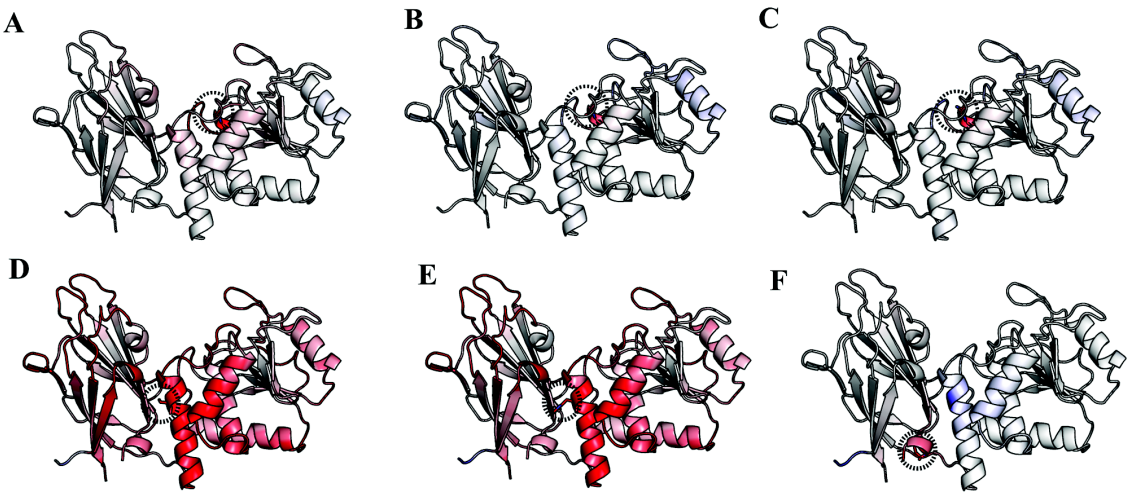
The results of the structure web-based tools are in kcal/mol and negative (ΔΔG) energy change indicates that the mutation is predicted to destabilize the mutant PTEN structure, while positive (ΔΔG) energy suggests stabilization of the structure.

CUPSAT was utilized to predict protein stability changes upon single amino acid point mutation using structural environment-specific atom potentials and torsion angle (phi, psi) potentials. The predicted unfavorable torsion angles for the p.G36E, p.C124S, p.R130G, p.R130Q, p.R173C, p.R173H, p.V191A and p.T348I mutations could be the result of the inability of the torsion angles (phi, psi) to adapt to their new environment which contributes to the higher impact of the atom potentials on stability, thus resulting in a stabilizing mutant. The p.R130G mutation results in both a destabilizing mutation and unfavorable torsion angles due to the greater extent of flexibility forcing the backbone into an incorrect conformation that disturbs the local structure.

Interestingly, the effect of these mutations on structural stability and dynamics can be further assessed with ENCoM in **Figure 3.6**. We chose ENCoM to predict the effect of mutations as it accounts for side-chain and long-range interactions offering a more realistic representation of intramolecular interactions. The p.R130G/L/Q, and p.R173C/H mutants have an increase in dynamics across the phosphatase-C2 domain interface. It has been previously posited that a

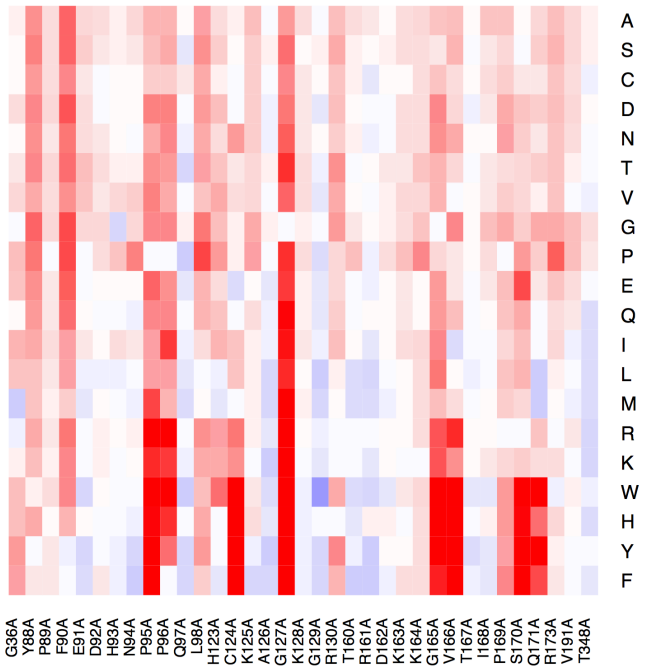
majority of mutations within the active-site pocket exhibit a loss of phosphatase activity as a result of the p.R173C/H mutations, further implicating this region as a possible mutation-driven allosteric interface. R173 forms a salt-bridge with D324 stabilizing the rich hydrogen bond network within the interface. Mutations located at position 173 result in a loss of charge and are involved in the destruction of this salt-bridge thus impairing the structural stability within the domain interface, contributing to a severe phenotype. This further highlights that the integrity of the interface is important for the overall stability of PTEN and these mutations could disrupt the global dynamics of the protein.

Figure 3.6 Effect of PTEN mutations on structural stability and dynamics



To identify the overall sequence tolerance or effect that other amino acids impose at the 13 identified positions, a heatmap was generated utilizing ENCoM, which calculates the ΔΔG distribution of the 20 essential amino acids for each of the given thirteen mutation positions (**Figure 3.7**). This heatmap allows the identification of hotspots throughout PTEN that affect structural stability. This is crucial as pathogenic mutations leading to monogenic diseases reveals a relationship between the destabilizing effects of mutations and severity of disease by as little as 1–3 kcal/mol. Additionally, since the evolution of a gain-in function or even a loss-in-function will be driven by a mutation of amino acids in key positions which may hinder interactions that ensure the stability of PTEN's fold to the degree that it renders the protein non-functional. Of the 13 mutations, positions G36, H123, C124, R130, and R173 possessed the largest stability decreases across the majority of the 20 amino acids. Conversely, positions G129, V191, and T348 indicated an increase in stability across half of the 20 amino acids. These data provide further understanding of the effect of mutations and the phenotypic consequences at these positions.

Figure 3.7 PTEN mutation heatmap



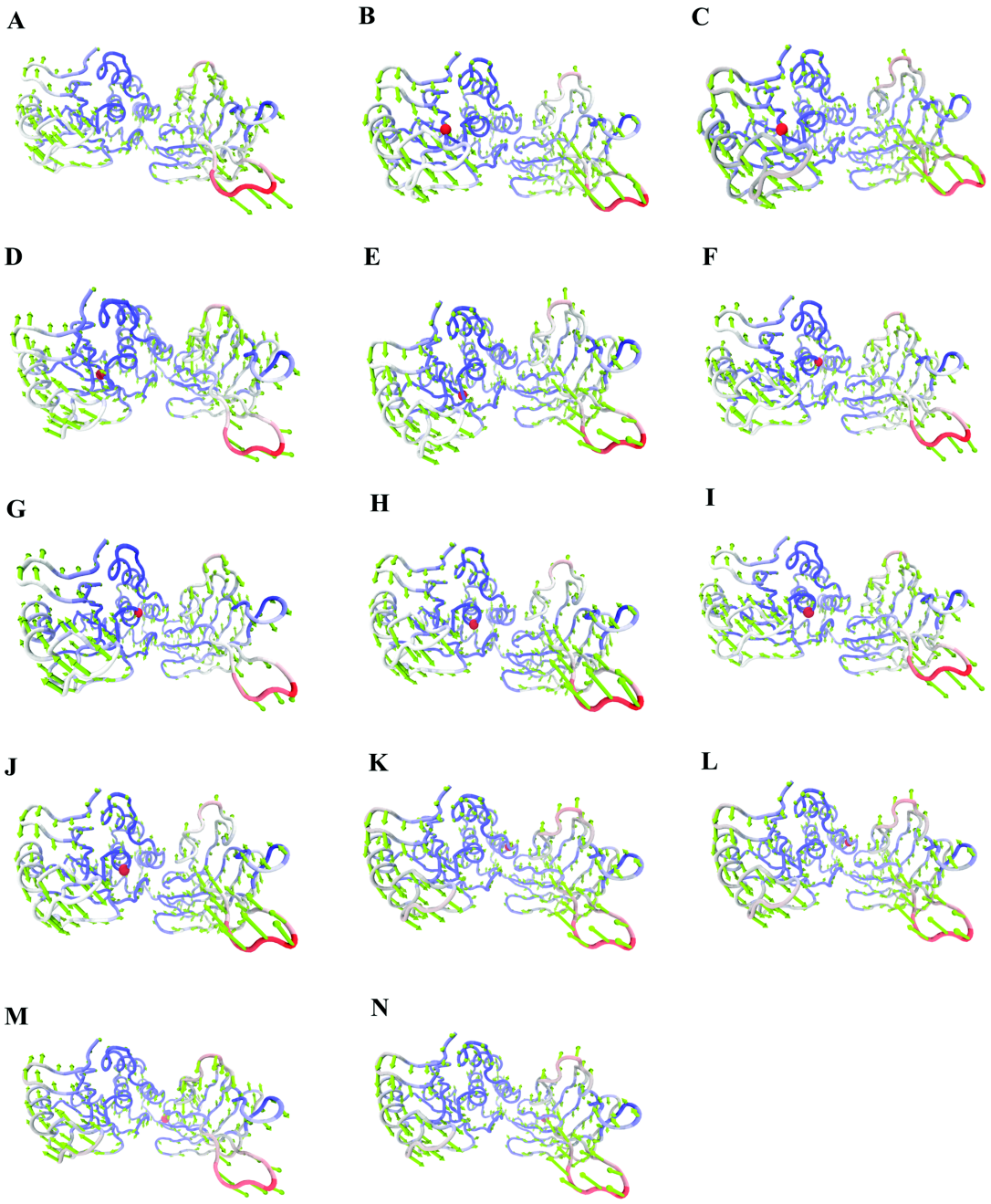
3.3.4 Correlated motions and effects of PTEN mutations on global dynamics

To understand how the intrinsic and global dynamics are impacted by each of the missense mutations on PTEN, elastic network models and normal mode analysis were utilized. To achieve this goal, we utilized the ProDy software package[287] for the construction of anisotropic network[225] coarse-grained models of both WT PTEN and mutants to further explore the interplay of the structural and dynamic effects each of the mutants have on the global dynamics of the tertiary structure. Correlation between collective modes was predicted by the calculation of the anisotropic normal mode (ANM) for each of the WT and mutant models. Lower frequency modes contribute to the covariance and entail the most cooperation and largest amplitude motions. ANM analysis exhibited a high cumulative involvement coefficient for the first 10 slow modes. The mode which emphasizes the contribution of the slowest modes and

contributes the most to the overall global motions of each model is mode 2 which has the lowest nonzero "frequency" and involvement coefficient of CI_2=1.00 best describing the conformational changes between the WT and mutant models.

Additionally, though pairwise cross-correlations of the residue fluctuations of the first lowest eigenvectors (first three modes) are quite similar, the P loop and WPD loop regions were the most flexible in ANM mode 2. Therefore, of the 10 slowest nontrivial normal modes that were analyzed, ANM mode 2 revealed the greatest differences in the motions of WT and mutant models of PTEN. These ANM 2 modes were found to describe variations of "hinge-bending" and "zipper-lie" global movements, thus characterizing their global conformational changes and cooperative motions relative to their mutation (**Figure 3.8A-N**).

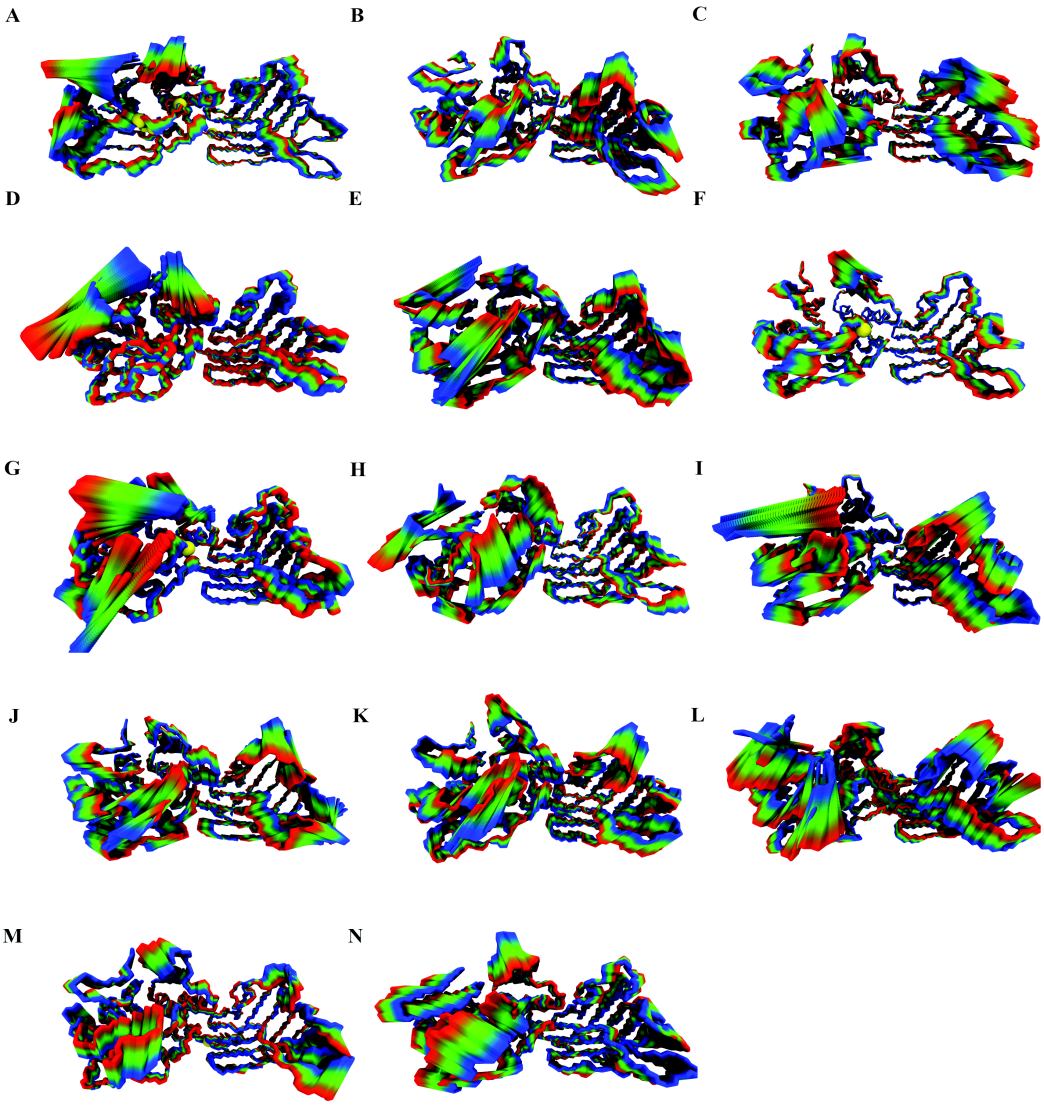
Figure 3.8 Global motions of WT PTEN and mutants



Furthermore, to better discriminate the effects induced by each of the mutations on the protein structure, an all-atom normal mode analysis was conducted by calculating principle component analysis on molecular dynamics trajectory ensembles of the WT and mutant models.

The projections of the displacement described by the first principle components on the simulated frames were analyzed. The first principal component of each WT and mutant model collectively account for more than ~40% of the total motion and is therefore a suitable approach to analyze protein dynamics. The introduction of a large charged residue (p.G36E/R) causes both local effects on the N-terminal region and the C2 domain (**Figures 3.9B,C**).

Figure 3.9 Protein dynamics fingerprint of WT PTEN and mutants



The p.H123Y, p.C124S, and p.G129E mutations (**Figures 3.9D-F**) significantly affect the principal motions of the TI loop and are somewhat close to the WT fingerprint, whereas the other mutations have well-defined dynamic effects. Additionally, the p.C124S mutation (**Figure 3.9E**) has dramatic effects on the N-terminal region and the catalytic loops as well as the CBR3 loop in the C2 domain indicating this mutation induces long-range perturbations. The p.G129R mutation (**Figure 3.9G**), though somewhat similar to p.G129E, has slightly more fluctuation to its principal motions most particularly in the N-terminal region and CBR3 loop. The p.R130G/L/Q mutations (**Figures 3.9H-J**) significantly alter PTEN dynamics in the surrounding active-site as well as long-range perturbations mostly due to the fact that the R130 position is catalytic.

Interestingly, p.R173C/H mutations (**Figures 3.9K,L**) affect the principal motions of the P, WPD, and TI loops in the active-site, and the inter-domain interface, illustrating long-range perturbations. The p.V191A mutation (**Figure 3.9M**) exhibits dynamic patterns similar to p.C124S and p.R130L indicating that this mutation affects the global dynamics of the TI loop in the active-site as well as the CBR3 loop. Lastly, T348I mutation (**Figure 3.9N**) reveals principal motions similar to H123Y indicating that this mutation greatly affects the global dynamics of the P, WPD, and TI loops in the active-site. Overall, each of the mutation motions captures the differences in the global dynamics induced by the effects of each mutation and reveal unique long-range perturbations within the PTEN protein.

3.3.5 Effects of ionization states on PTEN mutations

To understand whether the ionization states of titratable groups are affected by missense mutations, pKa calculations were predicted. pKa predictions provide insight into the local environment of critical residues within the active-site where the majority of the mutations are clustered (**Table 3.4**). The pKa predictions were performed on the apo form of PTEN using the 1D5R structure without the TLA substrate. Thus, the apo conformation exposes the active-site

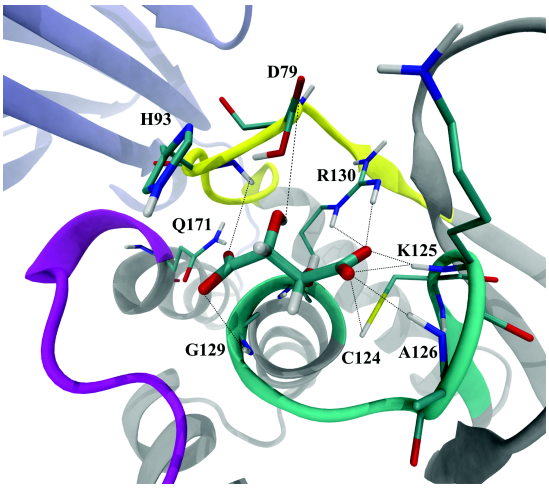
and was treated with a high dielectric constant of water. **Figure 3.10** illustrates interactions between WT PTEN ionizable residues and the tartrate molecule within the active-site. The pKa can shift depending on the surrounding environment thus affecting protein stability and activity.

Table 3.4 PTEN WT pKa calculations

Residue	pKa Expected	pKa Observed	pKa Delta
D92	4.00	-2.71	-6.72
H93	6.30	5.93	-0.37
D115	4.99	6.57	2.57
H123	6.30	4.59	-1.71
C124	8.30	3.79	-4.51
K125	10.50	>>pH	>>pH
K128	10.50	8.94	-1.56
R130	12.40	>>pH	>>pH

pKa calculations were performed with wild-type 1D5R structure without substrate. The pKa of the catalytic residue C124 showed a substantial shift toward greater acidity, which is attributable to the surrounding high density of basic residues.

Figure 3.10 PTEN active-site with tartrate molecule



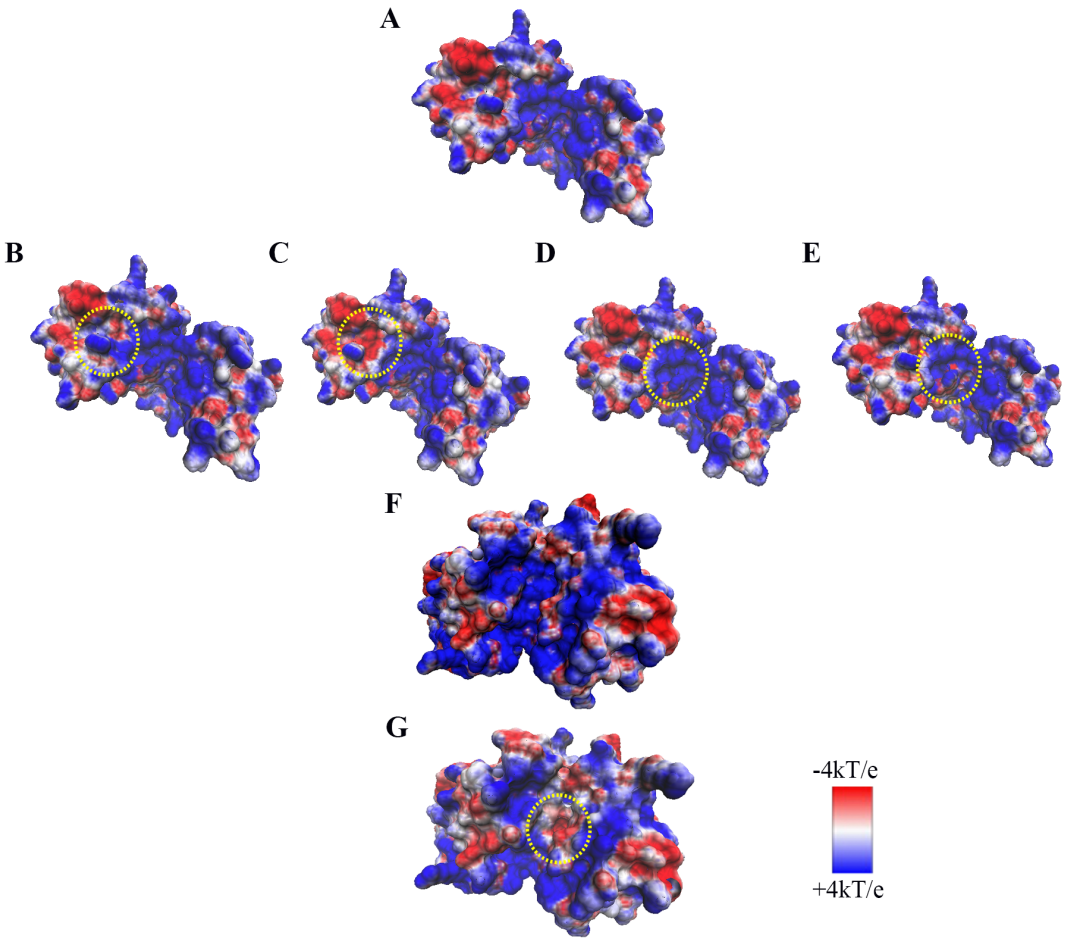
The pKa of D92 was substantially shifted toward greater acidity due to the formation of hydrogen bonds with H93 and K125, thus stabilizing the deprotonated state. H93 is critical for the

affinity of negative substrates thus a moderate shift of 5.93 is critical to maintain the catalytically competent protonation state of the residue. The elevated pKa value of D115 is attributable to hydrogen-bond formation with basic residues R130 and K125, serving as a hydrogen-bond acceptor. The pKa of H123 is slightly depressed due to hydrogen-bond formation with Y76, and disruption of this hydrogen bonding in the p.H123Y mutation leads to destabilization of the protein (**Table 3.3**). C124 exhibits a substantial shift toward greater acidity, which is attributable to the surrounding high density of basic residues H93, K125, K128, and R130 (**Figure 3.10**) indicating that cysteine is stabilized in its thiolate form at physiological pH. A mutation from cysteine to serine would disrupt the electrostatic environment thus further perturbing the pKa.

3.3.6 Electrostatic effects of PTEN mutations

Electrostatic surface potential calculations were generated for WT PTEN and mutants. As expected, WT PTEN (apo) is predicted to have a highly positive electrostatic potential surrounding the active-site (**Figure 3.11)**. Mutations within the active-site: p.H123Y, p.R130G, p.R130Q, and p.R130L reduce positive charge resulting in less favorable electrostatic, binding, and specificity interactions with PIP_3, thus contributing to a severe phenotype. The introduction of a large charged carboxylate side chain for buried mutations: p.G36R, p.G36E, p.G129R, and p.G129E imposes backbone conformational constraints, the inability to facilitate loop movement, and exposure to solvent. More specifically, the p.G129E mutation would thus experience electrostatic repulsion when binding to PIP_3 (**Figure 3.11**). The p.R173C/H mutations replace a large residue (R) with smaller ones (C/H), causing a loss in charge while disrupting the inter-domain hydrogen-bond network (**Figures 3.5C,D** and **3.11**).

Figure 3.11 Electrostatic surface potential of WT PTEN and mutants



Altogether, our observed predictions correlated with a decrease in protein structure stability in each of the associated mutations. I-Mutant 3.0 predicts all 13 mutations to be destabilizing. ENCoM predicts six mutations, (p.R130G/L/Q, p.R173C/H, and p.V191A) to be destabilizing. The effect of these mutations on structural stability and dynamics were further assessed in **Figures 3.6-3.9**. Our results indicate that the p.R130G/L/Q and p.R173C/H mutants have an increase in dynamics across the phosphatase-C2 domain interface. We predict that mutations at position 130 that result in a change in polarity or charge would result in a reduction

in structural stability, thus, likely impairing the ability to participate in the catalytic mechanism. It is possible that altering WT interaction of this active-site residue would affect PTEN function and thus lead to severe phenotype. The loss of positive charge at the R130 position (**Figure 3.11**) as well as interaction with neighboring essential catalytic residues C124 and D92 would disrupt overall function of PTEN.

The normal catalytic mechanism involves D92 donating a proton to the bridging oxygen of its substrate PIP_3.[27] Subsequently, R130 binds to the D3-phopsphate of PIP_3 and transfers it to neighboring C124; this interaction is possibly lost due to mutations at position 130.[27] Previous studies indicate that mutations within the active-site pocket, in addition to p.R173C/H mutants, exhibit inactivated phosphatase activity,[6,181,289,295] revealing a possible long-range communication, further implicating this region as a possible mutation-driven allosteric interface. R173 forms a salt-bridge with D324 stabilizing the rich hydrogen bond network within the interface. We predict that mutations at position 173 that result in a loss of positive charge would invoke the destruction of this salt-bridge thus likely to impair the structural stability within the domain interface leading to a severe phenotype. This further emphasizes that the integrity of the interface is important for the overall stability of PTEN and that these mutations could disrupt the global dynamics of the protein. In CUPSAT, eight mutations (p.G36R, p.H123Y, p.G129E, p.G129R, p.R130G, p.R173C, p.V191A, and p.T348I) are predicted to be destabilizing. This is in agreement with experimental data that suggests the loss of lipid phosphatase activity towards PIP_3 as a result of the loss of stability in the protein. The evolutionary-conservative analysis indicated that the majority of the disease-associated mutations are located in highly conserved parts of the PTEN structure (P loop as well as inter-domain region) (**Figure 3.4**). Our most notable observation was the identification of the inter-domain disruption caused by the mutations at positions 130 and 173 that lead to the most dramatic effect on PTEN long-range communication with the protein and its overall function implicating a rather interesting interplay between these two positions that may be

involved in a potential mutation-driven allosteric interface. This provides motivation for future studies to further investigate a network of interactions that contribute to a possible allosteric signal transmission and long-range communication. It is known that PTEN plays a key role as a tumor suppressor protein and is involved in a variety of cellular functions and protein-protein interactions.[296] Therefore, any significant changes in structural stability will modify the overall function of the protein having a global impact on various pathways involved in tumor suppression regulation.

Our results suggest that the replacement of cysteine with serine would greatly alter the dynamics of the N-terminal region, catalytic loops, as well as the CBR3 loop affecting membrane binding, indicating this mutation induces long-range perturbations (**Figure 3.9**). The p.G129R mutation introduces a larger, positively charged residue into a highly conserved region that destabilizes the structure and thereby contributes to a more rigid P loop (**Figure 3.9**). The flexibility of the native glycine residue may be necessary for PTEN's function. Additionally, mutations p.C124S, p.G129E, and p.R130G alter the polarity and charge of the active-site, leading to structural instability and a catalytically inactive mutant. Interestingly, recent experimental research implicates mutations p.C124S and p.G129E as contributing to a more severe phenotype, providing evidence that these inactive mutants disturb the function of co-expressed WT protein and perhaps aggravate normal cellular behavior.[297] Conversely, mutants retaining partial loss of function associate more frequently with a milder phenotype,[297] this may also explain the mild phenotype seen in p.G129R as well as in the diverse range in severity of phenotype (from mild to severe) as seen in multiple mutations at one position as in p.G36R/E, p.R130G/L/Q, and p.R173C/H as investigated in this study.

It is important to note that missense mutations at position 130 (p.R130G, p.R130L, p.R130Q) are located within the phosphatase domain of the active-site and participate in catalytic activity. Recently, p.R130G was identified as belonging to multivariate PTEN mutations

associated with endometrial cancer.[290,298] Our results demonstrate position 130 to be involved in a potential long-range communication pathway with position 173. Additionally, in comparison to WT PTEN, we observed changes in overall global dynamics due to each of the R130 mutations (**Figures 3.6-3.9**). In the coarse-grained normal mode analyses, the R130 mutants revealed a "hinge-like" motion compared to the "scissor-like" motion of the WT PTEN (**Figure 3.8**). Whereas in the all-atom normal mode analysis of the principal components on the molecular dynamic ensembles, we see not only local effects on the active-site induced by the mutations but also disruption of the inter-domain and long-range perturbations.

Previous studies reveal that the majority of the mutations within the active-site pocket exhibit a loss of phosphatase activity in addition to the p.R173C/H mutants located within the inter-domain interface (**Table 3.1**). Mutation p.R173C is experimentally predicted to reduce PTEN's catalytic activity towards PIP_3 substrate[181] and thus its function as well. In many instances mutations affecting catalytic sites or protein interaction sites in general, can be sensitive to residue mutations distant from the active-site.[299] Our results predict that both p.R173C and p.R173H destabilize the structure and disrupt the inter-domain interface (**Table 3.3**) (**Figures 3.5C-D and 3.6-3.9**). These mutations are located within a buried hydrophobic core and a rich hydrogen-bond network, given this region is also highly conserved and integral to the hydrogen-bond network, it further confirms the extreme importance of this interface for the PTEN function. Mutations at position R173 have been found mutated in a variety of cancers and is recognized as being among one of the eight most frequently mutated residues of PTEN.[4]

In fact, within the COSMIC database alone, PTEN mutations at positions 130 and 173 are distinctly prevalent in endometrial, ovarian, and central nervous system tumors with a high frequency of position 130 mutations in endometrial and ovarian tumors, whereas only 4% of central nervous system (CNS) tumors also occur at this position. Additionally, 18% of CNS tumors correspond to position 173 with high frequency. We observed that mutations

p.R130G/L/Q and p.R173C/H have an increase in dynamics across the phosphatase-C2 domain interface and lead to a more severe phenotype (**Figures 3.6-3.9**).

The missense mutations studied within this work allowed for perturbation of the WT properties of PTEN and provided a unique insight into two molecular mechanisms involving mutations at R130 and R173. Though genotype-phenotype correlations of these mutations are not solely attributable to a destabilization in structure, they are a major causative factor in endometriosis, endometrial cancer, and ovarian cancer as highlighted in this research.

3.4 Conclusions

Our results are not aimed at predicting the definitive value of the ΔΔG changes for each of the missense mutations but rather to gain insight into the stabilizing or destabilizing effects they have on the overall structure of PTEN. The majority of the somatic missense mutations investigated in this study are located within a contiguous conserved cluster (**Figure 3.2**) formed around the active-site and highly conserved P loop as well as the inter-domain region inducing structural instability and thus significant phenotypic changes. Ultimately, the primary goal was to investigate the genotype-phenotype relationship for each of the missense mutations presented in order that our results could be utilized to develop a more definitive diagnosis protocol and to this end, develop novel therapeutics for treatment of endometriosis, endometrial cancer, and ovarian cancer.

Chapter 4: Inter-domain communication pathway mechanism and correlated motions in endometriosis and cancer – associated PTEN mutations

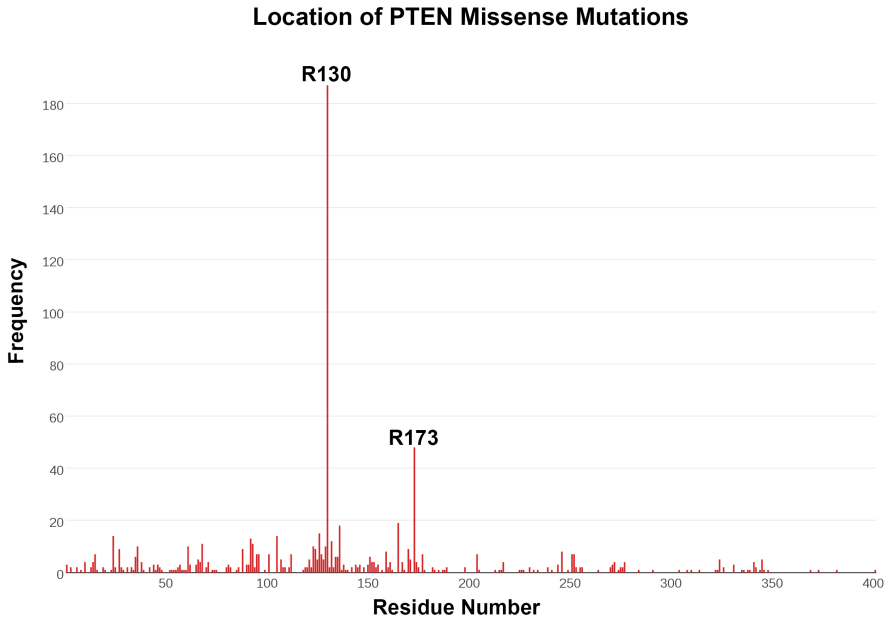
4.1 Introduction

Phosphatase and Tensin Homolog (PTEN) encodes a tumor suppressor protein that constitutes the main node of inhibition in the PI3K-Akt signaling pathway. Its most relevant function as a lipid phosphatase is based in its ability to catalyze the dephosphorylation of the 3' phosphate of the inositol ring in phosphatidylinositol (3,4,5)-triphosphate (PIP_3), an important intracellular lipid second messenger, thus restraining downstream anti-apoptotic and growth stimulatory effects that otherwise lead to unrestricted cell growth and tumorigenesis.[4,6,7] Alterations in the PI3K-Akt pathway give rise to various syndromes, metabolic dysfunctions, and cancer. We investigate the inter-domain communication pathway between both domains utilizing all-atom molecular dynamics (MD) simulations, communication pathway network analysis, and normal mode analysis. Communication network analysis was utilized to determine communities of residues that are critical to structural communication changes. Additionally, analysis of the residual local frustration and normal mode analysis of apo-and active-site bound (PIP_3-PTEN) are utilized to discern correlated motions associated with wild-type and cancer-associated PTEN mutations. We find that certain residues play a critical role in controlling salient inter-domain communication pathways and motions that contribute to both subtle structural changes as well as global correlated motions that are essential to the function of PTEN. The results elucidate a network mechanism that involves correlated loop motions that affect protein dynamics offering novel insights into the potential correlation of endometriosis and cancer-associated PTEN

mutations. PTEN has recently been implicated in the malignant transformation of endometriosis, endometrial, and ovarian cancers.[50,127,130,144-156]

Endometriosis is a gynecological disease that affects over 176 million women worldwide and is defined as the presentation of endometrial tissue outside (ectopic) the uterine cavity.[121] PTEN somatic missense mutations and deletions have been reported in endometriosis, endometrial and ovarian cancers.[147,156,176,182-189,276] Clinically, endometriosis has been identified in ~30% of cases with synchronous endometrioid type endometrial and ovarian cancers.[155,162-164] The inactivation of *PTEN* is an early event in the premalignant lesion known as endometrial hyperplasia, which leads to the development of ovarian and endometrial cancers.[169,176,191] Thus, somatic missense mutations may contribute to the genesis and development of endometriosis and subsequently cancer. The somatic missense mutations investigated here within, result from the analysis of our recent work identifying two potentially allosterically-driven mutation positions involved in all three of the aforementioned phenotypes.[300] These two positions are also frequently mutated in PTEN in various disorders and cancers such as glioblastoma,[301-304] Autism Spectrum Disorder,[295] and Anaplastic Astrocytoma.[305] **Figure 4.1** demonstrates the distribution and location of PTEN somatic missense mutations retrieved from the COSMIC database.[278] Therefore, a greater understanding of the effects of these mutations will offer insight into the disease mechanism and malignant transformation of endometriosis.

Figure 4.1 Location and frequency of PTEN missense mutations

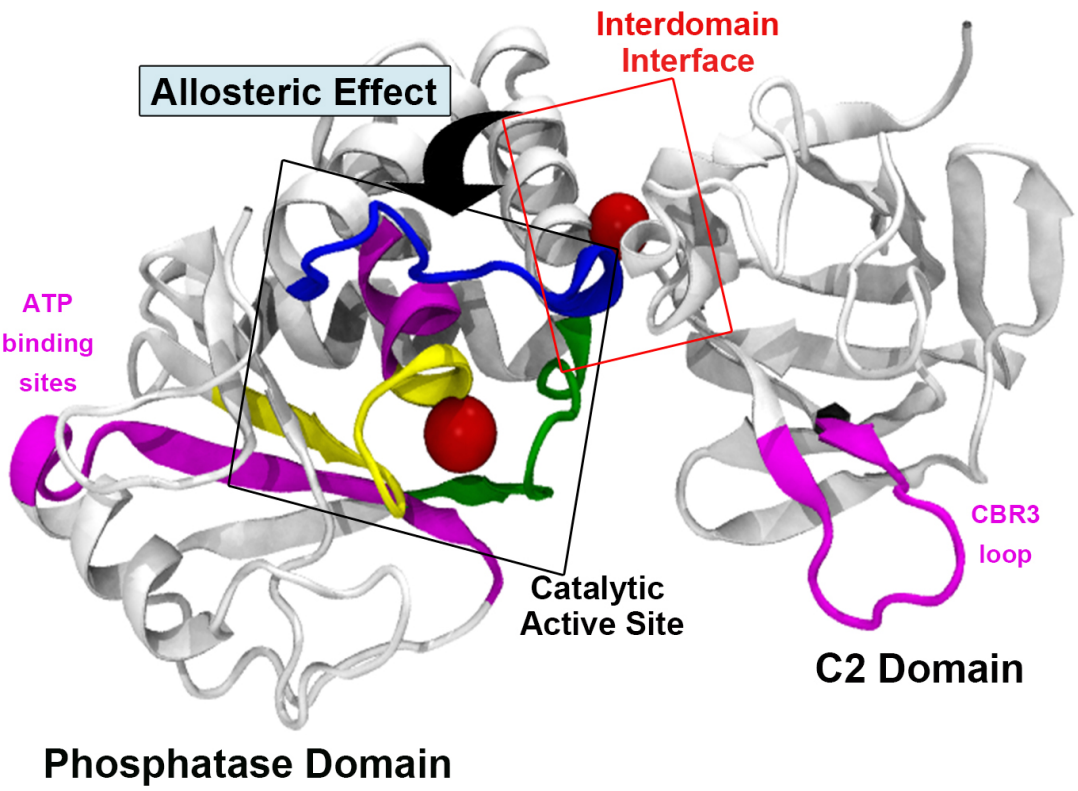


Structurally, PTEN consists of 403 amino acid residues which comprise two main domains: an N-Terminal phosphatase domain containing a central five-stranded β-sheet with six α-helices (two α helices on one side and four on another) wherein the active-site is delimited in part by three critical loops (P loop, WPD loop and TI loop) and a C-Terminal C2 domain that regulates membrane-binding which contains a β-sandwich that consists of two antiparallel β-sheets with two short α helices intervening between the strands.[4] The phosphatase and C2 domains associate across an extensive inter-domain interface that is adjacent to the active-site and consists of conserved residues frequently mutated in cancer.[4] The interactions of these two domains are controlled by three loops located in the catalytic active-site pocket, two (the P loop and WPD loop) of which share homology with both tensin and auxilin proteins that are commonly found in the regulatory sites of PTPs and DSPs. The phosphatase domain however contains a rather wide and deep catalytic pocket specific only to PTEN and is ~8 Å deep with an elliptical opening of ~5 x 11 Å.[4] The extension and larger width of the active-site pocket is due to

the four-residue insertion of the TI loop. The flexibility of the WPD loop is required to adopt a closed conformation bringing the catalytic Asp92 closer to PIP_3 enabling catalysis. The WPD loop also plays a similar role in the Yersinia protein tyrosine phosphatase (YopH) which is essential to the dynamics of the active-site.[306] Mutations in the TI loop (T167), WPD loop (H93 and D92), and P loop (C124, K128, G129, and R130) lead to a reduction in PIP_3 phosphatase activity by ~75%.[4]

Two additional motifs that also contribute to PTEN's function and stability include: (i) the CBR3 loop (residues 260-269), and (ii) two ATP-binding motifs (residues 60-73 and 122-136) the latter of which overlaps the P loop. The CBR3 loop facilitates membrane binding, which results in conformational changes that occur in the active-site leading to the assumption of an allosteric-activation mechanism in PTEN.[293,307-310] Interestingly, previous studies have revealed that a slight conformational change in PTEN exposes both ATP-binding regions; however, a more substantial conformational change would essentially preclude its activity as a phosphatase establishing the existence of a role not previously described for this tumor suppressor.[9,10,311] Moreover, a more recent study utilizing neutron scattering combined with all-atom MD simulations revealed the importance of the CBR3 loop in membrane and active-site interactions with phosphatidylinositide substrates ($PI(4,5)P_2$ and $PI(3,4,5)P_3$).[312] These inositolphosphates bind stronger to the protein in the absence of PIPs bound within the membrane.[312] Additionally, a cluster of contacts of accumulated $PI(4,5)P_2$ was observed at the C2 domain, which populated the CBR3 loop and the cleft (inter-domain) between the phosphatase and C2 domains (**Figure 4.2**). These results further implicate the importance of the CBR3 loop in the inter-domain dynamics as well as the communication pathway.

Figure 4.2 Structure of potential PTEN allosteric inter-domain area



Recently we identified an inter-domain disruption within PTEN caused by mutations at positions R130 and R173 with an increase in dynamics across the phosphatase-C2 domain interface.[300] Mutations at these two positions illustrated the most dramatic effect on PTEN long-range communication within the protein suggesting an interesting interplay of their involvement in a potential mutation-driven allosteric interface (**Figure 4.2**). The combined structural and normal mode analyses led us to hypothesize that the inter-domain interface allows allosteric regulation between the phosphatase and C2 domain via global conformational changes among the mutant residues; therefore linking the inter-domain interface with the phosphatase active-site. This offers fresh insight into an association between PTEN long-range communication and inter-domain allostery. Though traditionally, allostery has been defined as occurring between subunits in a multi-subunit system driven by a conformational change,[313] a new concept of allostery has

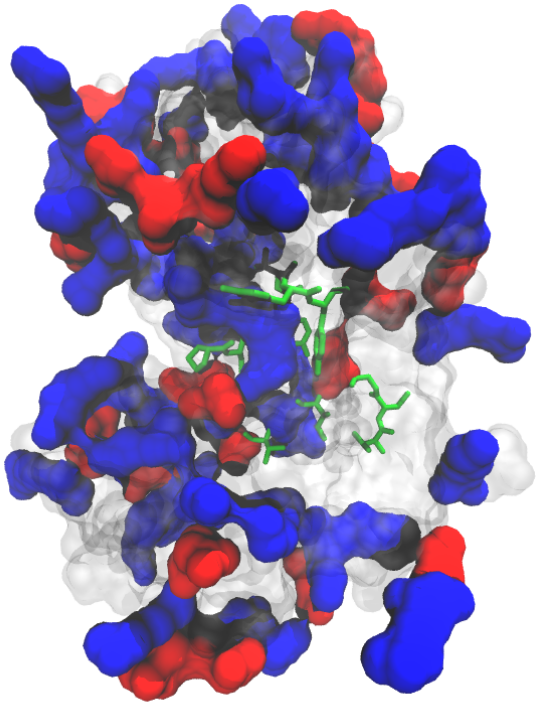
been established suggesting that it may occur without a conformational change,[314-316] and that all proteins may be capable of being allosterically regulated.

Despite the specific nature of PTEN's functional dynamics, the intra- and inter-protein conformational changes that govern its potential allosteric signal propagation have yet to be investigated and thoroughly defined. Moreover, missense mutations may have underlying effects on communication pathways that govern global conformational dynamics,[317,318] inter-domain destabilization effects,[319,320] propagation of allosteric signals,[318] and correlated mutations involved in pathways of allosteric communication,[318,321] all of which have yet to be identified within PTEN. Interestingly, missense mutations often occur at sites having a high local or global centrality in the protein structure network which are enriched by three-fold at the interaction interfaces of proteins associated with the various diseases.[322]

Our previous findings set the stage for the present work, which is a more direct investigation of PTEN inter-domain communication. In particular, a necessary hypothesis, for our mutation-driven allosteric hypothesis is that the missense mutations associated with endometriosis, endometrial, and ovarian cancers, should alter the internal atomic properties of the active-site and inter-domain interface pertinent to binding. To investigate this, we combined all-atom MD, PSN-ENM-NMA, and analysis of residual local frustration to further explore the effects missense mutations have on dynamics of PTEN with a particular interest in changes that occur within the active-site and inter-domain interface. Specifically, we have generated *in silico* R130G, R130L, R130Q, R173C, and R173H mutants. This combined approach allows us to analyze the long-range communication within PTEN and associated mutants of the endometriosis, endometrial, and ovarian cancer phenotypes. Our choice of the mutants, derives from the high frequency association of endometriosis and cancer at positions 130 and 173 with each of the aforementioned phenotypes as well as our previous structural studies of PTEN.[300] As indicated in the PTEN crystal structure (PDB ID 1D5R), mutations R130G/L/Q, are located within the P loop

of the active-site, adjoining the inter-domain interface where mutations R173C/H reside within the extensive hydrogen-bonding network (**Figure 4.2** and **4.3**).[4]

Figure 4.3 PTEN inter-domain interface



Inter-domain interface of PTEN with the interface residues highlighted in green licorice.

4.2 Computational methods

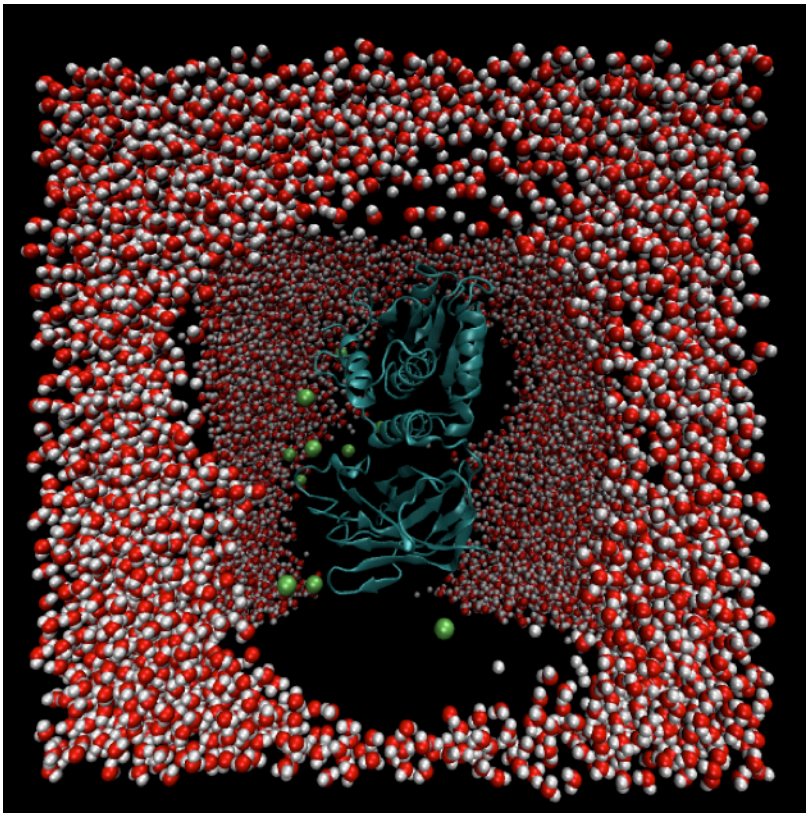
4.2.1 Molecular dynamics (MD) simulations

4.2.1.1 System set-up

The 2.1 Å X-ray crystal structure atomic coordinates for wild-type (WT) PTEN were obtained from the protein data bank [PDB ID: 1D5R].[4] The substrate-like tartrate (TLA) molecule was removed from the active-site for the apo-PTEN and mutant PTEN calculations. All simulations were performed using GROMCAS 4.6.3[270] with GROMOS53a6 force field[257]. Each

system was solvated inside a cubic box of simple point charge (SPC)[264] water with at least 10 Å of water between the protein and edges of the box. All simulations were performed in explicit solvent, with chloride (Cl-) counter ions added to obtain neutrality of the system. Periodic Boundary Conditions and a 2 fs time step were employed for each simulation. Each system contained roughly 75,825 atoms. The Particle Mesh Ewald (PME) method[288] was used to treat long-range electrostatic interactions and a cutoff of 9 Å was used for non-bonded interactions.

Figure 4.4 Representation of PTEN simulation box with water molecules and chloride ions



4.2.1.2 Energy minimization

A series of five overall minimization stages were performed on each PTEN model to remove any steric clashes and minimize the forces that were introduced as a result of the mutation

that was introduced to the WT structure. Each system was then subjected to a step-wise energy minimization using the steepest descent method.

4.2.1.3 Equilibration and production simulation

The minimized structures were then slowly heated from 0 to 300 K over 100 ps and equilibrated for an additional 250 ps. The system was heated to 300 K by linearly by increasing the temperature, through velocity rescaling, every 10 ps. Harmonic position restraints were applied to all heavy protein atoms in x, y, z dimensions (fx, fy, fz) starting from 1000 kJ/mol/nm^2 and gradually releasing to 0 kJ/mol/nm^2. The production runs were performed in the NPT (isobaric-isothermal) ensemble at 300 K. Bond lengths were constrained using LINear constraint solver (LINCS) algorithm[261] and the van der Waals forces were maintained at 1.4 nm. The Berendsen[266] weak coupling method was employed to maintain constant temperature with a temperature-coupling relaxation time of 0.1 ps, a pressure-coupling constant of 0.5 ps, and a compressibility of $4.5 \bullet 10^{-5}$. The total simulation time for each model was 200 ns and coordinates were recorded every 1 ps.

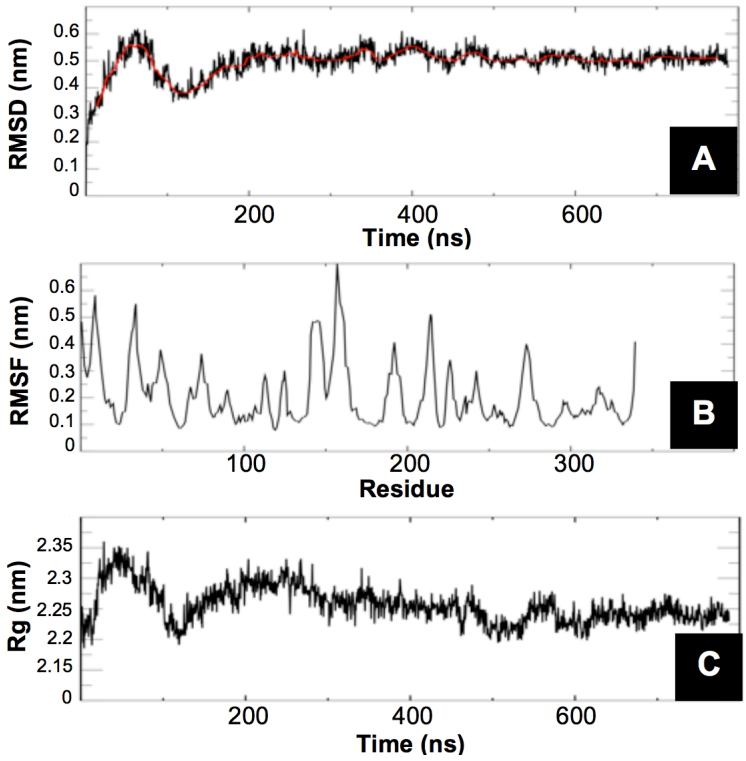
Table 4.1 PTEN MD simulations

System	WT	C124S	G129E	G129R	R130G	R173C	R173H
CPUs	240 (146 ns/day)	128 (113 ns/day)	128 (114 ns/day)	128 (105 ns/day)	128 (117 ns/day)	128 (112 ns/day)	128 (116 ns/day)
Residues	307	307	307	307	307	307	307
Water Molecules	24,139 (72,417 atoms)	24,140 (72,420 atoms)	24,138 (72,428 atoms)	24,138 (72,414 atoms)	24,142 (72,426 atoms)	24,137 (72,411 atoms)	24,143 (72,429 atoms)
Total Atoms	75,829	75,830	75,837	75,837	75,823	75,811	75,835
Total Cl- Atoms	12	11	10	12	10	10	10
Total Net Charge	+12	+11	+10	+12	+10	+10	+10
EM (Steepest Descent)	until convergence	until convergence	until convergence	until convergence	until convergence	until convergence	until convergence
EQ NVT	200 ps	200 ps	200 ps	200 ps	200 ps	200 ps	200 ps
EQ NPT	150 ps	150 ps	150 ps	150 ps	150 ps	150 ps	150 ps
Production Run	787 ns	200 ns	200 ns	200 ns	200 ns	200 ns	200 ns

System	G36E	G36R	H123Y	R130L	R130Q	V191A	T348I
CPUs	128 (111 ns/day)	128 (117 ns/day)	128 (116 ns/day)	128 (117 ns/day)	128 (114 ns/day)	128 (106 ns/day)	128 (116 ns/day)
Residues	307	307	307	307	307	307	307
Water Molecules	24,133 (72,339 atoms)	24,130 (72,390 atoms)	24,135 (72,405 atoms)	24,139 (72,417 atoms)	24,140 (72,420 atoms)	24,136 (72,408 atoms)	24,138 (72,414 atoms)
Total Atoms	75,813	75,813	75,819	75,818	75,824	75,816	75,824
Total Cl- Atoms	10	12	11	10	10	11	11
Total Net Charge	+10	+12	+11	+10	+10	+11	+11
EM (Steepest Descent)	until convergence	until convergence	until convergence	until convergence	until convergence	until convergence	until convergence
EQ NVT	200 ps	200 ps	200 ps	200 ps	200 ps	200 ps	200 ps
EQ NPT	150 ps	150 ps	150 ps	150 ps	150 ps	150 ps	150 ps
Production Run	200 ns	200 ns	200 ns	200 ns	200 ns	200 ns	200 ns

4.2.2 Convergence and analysis

Figure 4.5 Convergence analysis of WT PTEN



4.2.2.1 Root-mean-square deviation (RMSD)

The production phase of each simulation was monitored by checking the system temperature at 300 K which remained fairly constant throughout the simulations. The RMSD of the backbone was calculated with **g_rmsd** a module within the GROMACS program (**see figure 4.5A**).

4.2.2.2 Root-mean-square fluctuation (RMSF)

The fluctuation of the individual residues was calculated using **g_rmsf** a module within the GROMACS program (**see figure 4.5B**).

4.2.2.3 Radius of gyration (Rg)

The compactness of the overall protein was calculated using **g_gyrate** a module within the GROMACS program (**see figure 4.5C**).

4.2.3 Protein structure network (PSN) communication pathway and community network analysis

Long-range communication and allosteric networks were characterized using a mixed protein structure network (PSN) and elastic network model-normal mode analysis (ENM-NMA) approach previously applied to investigate structural and allosteric communication pathways.[236,237,323,324] Based on early work established by Vishveshwara *et al.*,[234,235] the PSN is constructed from the atomic coordinates of residues, which represent the nodes of the network. Two nodes are connected by an edge if the percentage of the interaction between them is greater than or equal to a given Interaction Strength cut-off:

$$I_{ij} = \frac{n_{ij}}{\sqrt{N_{ij}N_{ij}}}100 \qquad \text{(Equation 4.1 } \textit{Interaction Strength cut-off}\text{)}$$

where $\boldsymbol{I_{ij}}$ is the interaction percentage of nodes $\boldsymbol{i}$ and $\boldsymbol{j}$, is the number of side-chain atom pairs within a given cut-off (4.5 Å), and $\boldsymbol{N_i}$ and $\boldsymbol{N_j}$ are, respectively, the normalization factors (NF) for residues $\boldsymbol{i}$ and $\boldsymbol{j}$, which take into account the difference in size of different nodes and their propensity to make the maximum number of contacts with other nodes in protein structures.

The mixed PSN-ENM approach was recently implemented to predict salient structural communication pathway in biomolecular systems. The underpinning of WebPSN[236] lies with a Protein Structure Graph (PSG) and searches for all shortest communication pathways between user-specified residues. A PSG defines amino acids as nodes and the non-covalent interactions

among them as links. Such graphs are useful in identifying clusters of residues that stabilize the protein structure and protein-protein interfaces.[234] The network topology of PSGs depends on the cut-off of the interaction strength between residues used in the constructed graph.

The mixed PSN-ENM method (WebPSN) involves a multi-step process where network features (i.e. nodes, hubs, links, etc.) are computed by building a PSG and the shortest communication pathways on ensembles of structures are acquired from a single high-resolution structure. The algorithm defines all possible communication paths between selected node pairs and filters the results to cross-correlation of atomic motions, as derived from ENM-NM. Filtering consists in retaining only the shortest path(s) that contains at least one residue correlated (i.e. with a cross-correlation value ≥0.6) with either one of the two extremes (i.e. the first and last residues in the path). Meta-paths made of the most recurrent nodes and links in the path pool (i.e. global meta-paths) and infer a coarse/global picture of the structural communication in the considered system. In detail, meta-paths are made of nodes ≥5% of the considered path pool (i.e. 'frequent nodes'), and of links satisfying both conditions of being present in of the paths and of connecting 'frequent nodes'.

4.2.4 Residue perturbation – local frustration analysis

A frustratometer algorithm was developed by the Wolynes and Komives group to determine residual local frustration that measures whether contact between residues is energetically optimized or not and evaluates how such changes affect the interaction energies.[325] Residues that are systematically destabilized by random changes in their vicinities are considered minimally frustrated, while those that are systemically stabilized are considered highly frustrated.[325] To quantify local frustration in WT PTEN (PDB ID 1D5R), PTEN mutants, and PTEN-PIP_3 complex, we utilized the frustratometer webserver[326] (http://frustratometer.qb.fcen.uba.ar) to assess the local frustration. For this work we utilized,

"mutational frustration" index, in which the decoy set involves randomizing the identities of the interacting amino acids, keeping all other interaction parameters at their native value. This scheme effectively evaluates every possible mutation of the amino acid pair that forms a particular contact in a robustly fixed structure. Minimally frustrated contacts in green and highly frustrated contacts in red are depicted of the contact on the structural models. The direct inter-residue interactions are illustrated with solid lines and the water-mediated interactions with dashed lines. Projection of local frustration of the average frustration scores overall the contacts made by a particular residue are also plotted per-residue with the number of contact within 5 Å of the Cα of each residue.

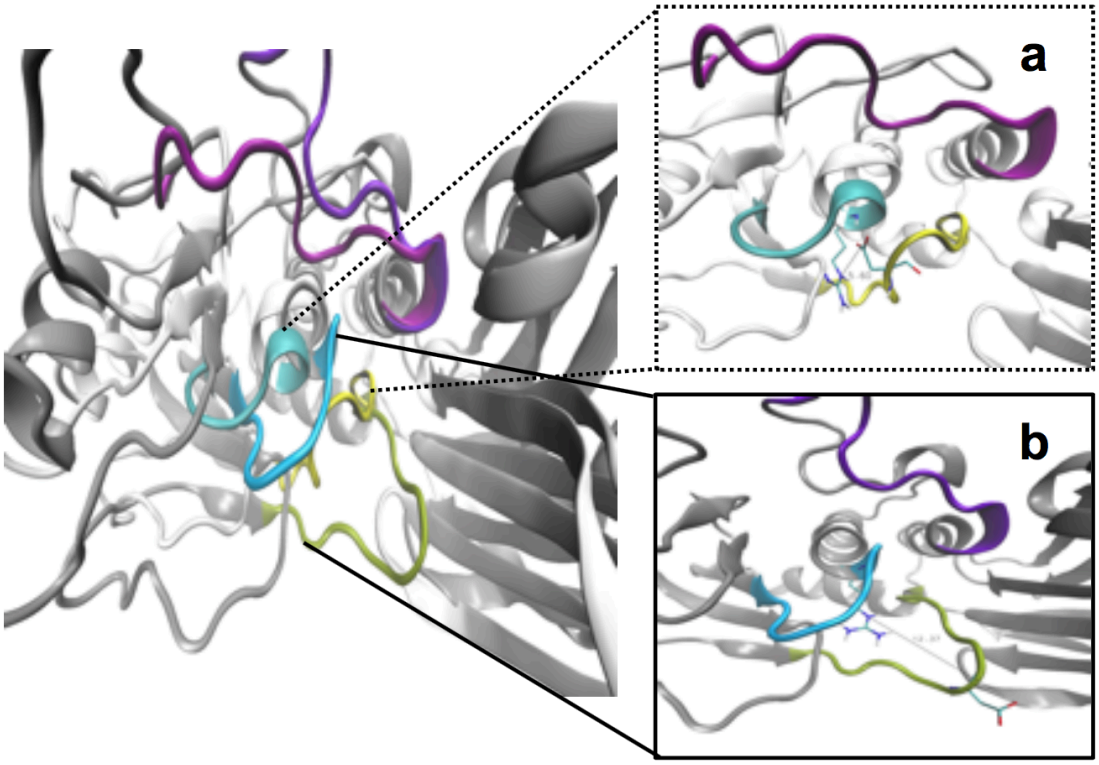
4.3 Results and discussion

4.3.1 The effect of mutations on PTEN structure

To observe the dynamical behavior of residues, the RMSF value was calculated in the WT PTEN system from each Cα atom revealing dramatic fluctuation within the range of ~0.1-0.58 nm in the n-terminal region and the WPD (residues 88-98) and P (residues 128-130) active-site loops **Figure 4.5B**. The TI loop (residues 160-171) attained maximum level of fluctuation up to ~0.7 nm. To access the intra molecular compactness of the overall protein, the radius of gyration (Rg) was calculated over the lifetime of the simulation. There was an increase in the first 50 ns and a subsequent decrease at ~128 ns prior to convergence (**Figure 4.5C)**. Though these changes were minor and within a range of less than 0.15 nm, a closer inspection revealed the destruction of a critical salt-bridge involving catalytic residues R130 and D92. Further analysis revealed dramatic changes in the active-site loop conformation. **Figure 4.6** illustrates the considerable deviation in the distance between R130 and D92. At 50 ns the distance between R130 NE2-D92 OD1 is 5.60 Å, however at 128 ns there is a disparate change in distance (12.37

Å). It appears that the WPD loop is initially in a "closed" conformation, strategically positioning the catalytically important aspartate residue closing over the active-site. However, the large conformational change destroys the R130-D92 salt-bridge moving the aspartate far out of the active-site representing an "open" conformation.

Figure 4.6 Conformation change – "closed" to "open"



4.3.2 Inter-domain communication pathway

4.3.2.1 Inter-domain meta-path analysis

To specifically identify a relevant path of communication within the structure of each PTEN system, a mixed protein structure network (PSN) and elastic network model (ENM) approach was employed.[236,237,323,324] The building of the protein structure graph (PSG) was carried

out by means of the PSN algorithm. The strength of interaction between residues *i* and *j* (I_{ij}) was evaluated as a percentage, then I_{ij} was calculated for all node pairs. The interaction strength, I_{min} was chosen and any residue pair for which $I_{ij} \geq I_{min}$ is considered to be interacting and hence connected to the PSG. The network components and parameters for each system are outlined in **Table 4.2.**

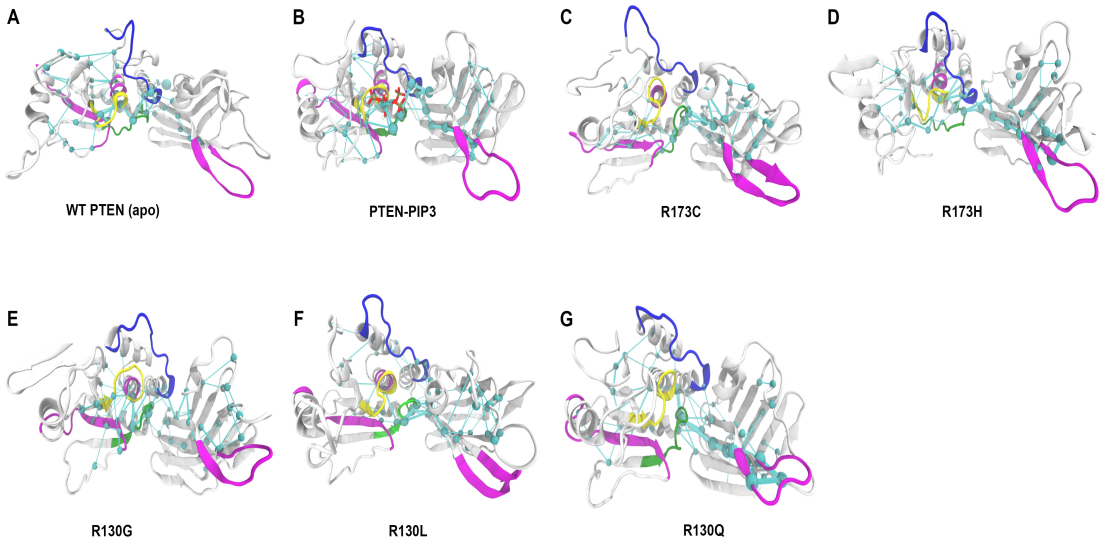
Table 4.2 Network components and parameters

	PTEN (apo)	**PTEN-PIP_3**	**R130G**	**R130L**	**R130Q**	**R173C**	**R173H**
I_{min}	3.36	3.97	4.11	3.51	3.38	3.89	3.5
Hubs	35	37	26	36	50	28	30
Links	226	219	166	217	255	193	221
Paths	11302	26609	12015	15267	37935	14092	18434
Nodes	193	189	139	185	193	168	179

In a previous study, we calculated anisotropic network analysis (ANM) and all-atom principal component analysis (PCA) of atomic fluctuations which highlighted "hinge-bending" and "zipper-like" PTEN motions as well as the dynamic fingerprint and displacement of the active-site loops, CBR3 loop, and inter-domain interface,[300] to which these inferences lend their support in ascertaining the underlying structural communication pathway of PTEN. The PTEN mutations, PTEN-PIP_3 substrate interactions as well as conformational dynamics, all exert differential impacts on structural communication. The latter can be examined by representing PTEN structures as an elastic network of interacting residues. PSN analysis was utilized to deduce the effects mutations have on the native fingerprint and stable residues within the structure network.

4.3.2.1.1 Meta-path analysis of WT PTEN

Figure 4.7 Meta-path analysis

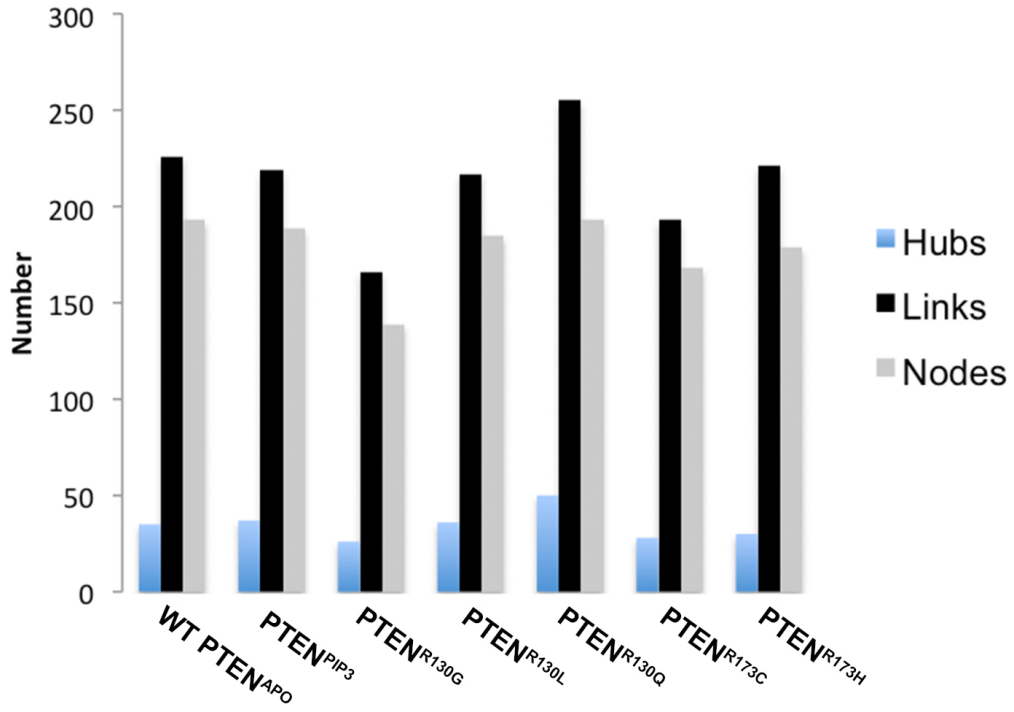


A detailed comparison of the communication paths in PTEN allows us to identify distinct long-range salient communication pathways. **Figure 4.7** highlights the meta-path, of each PTEN system, which contains regions that participate in the stability of the protein core (inter-domain) in the overall dynamics of the protein. Our results reveal that though the WT PTEN (apo) has an identifiable communication pathway that is more spread out over the N-terminal region of the protein structure, it's strongest interaction is within the inter-domain interface (**Figure 4.7A**). The PTEN-PIP_3 complex has an even more distinguishable pathway that illustrates the strongest interaction to be in the WPD loop (residues 88-98) of the active-site that moves into the core of the protein and spreading throughout the C2 domain (**Figure 4.7B**). Interestingly, as expected, upon ligand binding the phosphatase domain acquires hubs in the active-site as well as inter-domain region.

4.3.2.1.2 Meta-path of cancer-associated PTEN mutants

The remaining meta-paths involving the PTEN mutants captured a rather significant scheme in their pathways of communication. Each of the mutants had the active-site loops involved in their pathways most particularly the WPD loop (residues 88-98) which contains the acidic/basic D92 that participates in catalysis. They share several other features that consist of residues in their hubs containing both ATP-binding sites (residues 60-73 and 122-136) in addition to the CBR3 loop (residues 260-269). However, $PTEN^{R130Q}$ has only a few active-site hubs in its pathway (**Figure 4.7G**). It is known that various misfolding mutations tend to impair selective native stable hubs; the extent to which this structural effect appears is related to the extent of the biochemical defect associated with the mutation.[234] Interestingly, $PTEN^{R130G}$ demonstrates a more clustered pathway concentrated in its core that affects the compactness of the protein structure (**Figure 4.7E**). A recent study revealed that $PTEN^{R130G}$ mutant alters the polarity and charge of the active-site leading to a structurally instable and catalytically inactive mutant.[327] Consistent with the catalytically active mutants $PTEN^{R130L}$, $PTEN^{R130Q}$, $PTEN^{R173C}$, and $PTEN^{R173C}$, the number of hubs in the WT PTEN (apo) and PTEN-PIP_3 complex systems were higher than the catalytically inactive $PTEN^{R130G}$ which had lowest number of hubs as seen in **Table 4.2** and **Figure 4.7**.

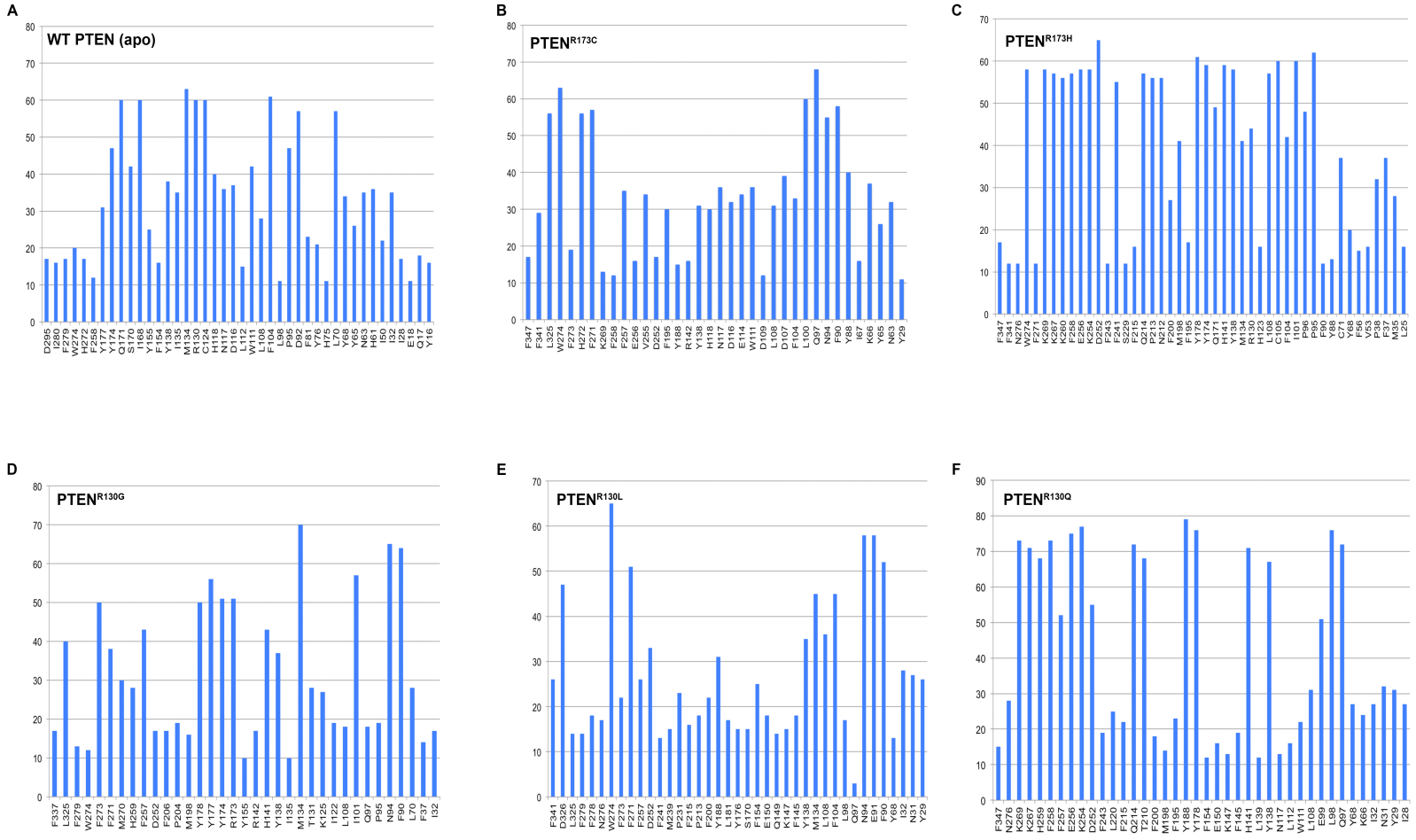
Figure 4.8 Nodes, hubs, and links



This further emphasizes the role of the highly conserved R130 position both in structural stability, and intra-protein allosteric communication, as it is an integral hub residue participating in both inactive and active states. The PTEN^{R130Q} mutant is richer overall in hubs, links, and nodes. Moreover, its communication pathway obtained highly enriched hubs within inter-domain region as well as the CBR3 loop (residue 260-269) (**Figure 4.7G**). Similarly, PTEN^{R130Q} demonstrated the greatest frequency of nodes (percentage) overall in the whole pool of paths (**Figure 4.9**). The PTEN-PIP_3 complex was third overall with the highest hubs, links and nodes; yet it revealed the most distinctive inter-domain communication pathway. Though the R173 mutant systems didn't reveal a distinct communication pathway, they had the fewest hubs second only to PTEN^{R130G} (**Table 4.2** and **Figure 4.8**). This can be attributed to the mutation being in the

central core of the protein; therefore decreasing normal interactions required for their communication network.

Figure 4.9 Frequency of nodes in the whole pool paths



Several active-site residues, ATP-binding site resides and CBR3 loop residues were found to participate in the frequency of nodes of all systems in the whole pool of paths that played a critical role in the salient communication pathway (**Figure 4.9**). In comparing the WT PTEN (apo) system with the mutants, the WT demonstrated several residues that had a high frequency of nodes within its pool of paths. Most particularly, the residues belonged in the CBR3 loop (residues W261, F266, and I267), TI loop (residues Y161 and Y164), WPD loop (F91, L95, and W98), ATP-binding site (H62, Y63, and F68), and the critical and highly conserved R130 residue (with 60% frequency) in the central node of the active-site.

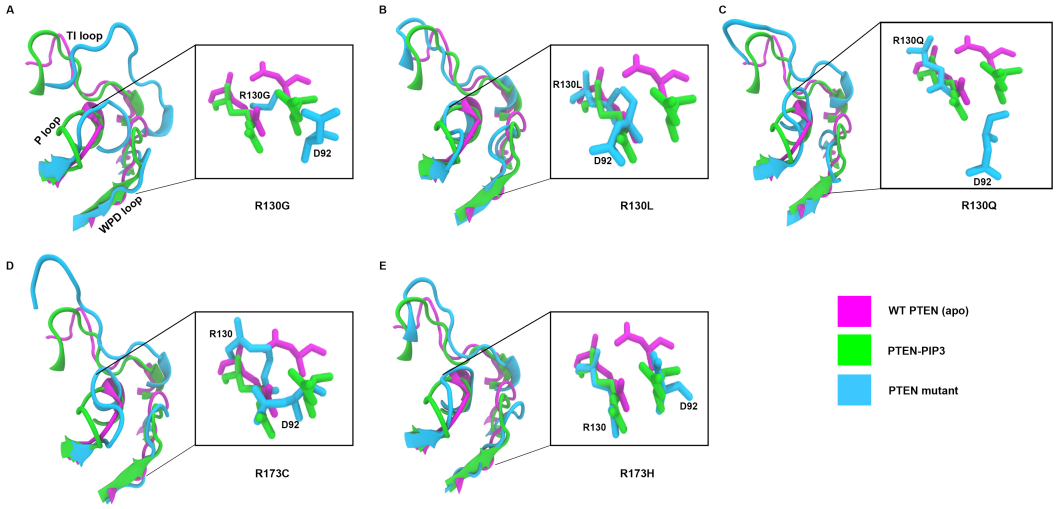
$PTEN^{R130G}$ only had a few residues that belong to the active-site as part of the frequency of nodes within its pool of paths such as the WPD loop (residues F90, N93, P95, and Q97), P loop (residues T131, M134, and I135) and the conserved R173 residue (with 50% frequency) location within the inter-domain interface. $PTEN^{R130L}$ only had residues in the WPD loop (residues F90, E91, N94, Q97, and L98) appear in its pool of paths. Interestingly, Q97 had a frequency of 75%, which is adjacent to the acidic D92 that interacts with R130 in catalysis. Additionally, W274 appears with a 65% frequency, which is a rather significant as this residue participates in hydrogen bonding with R173 within the rich hydrogen-bond network of the inter-domain interface. In the R173C/H mutations, steric clash is imposed and side chain distortion occurs abolishing the W274 and R173 hydrogen-bond interaction. Similarly, $PTEN^{R130Q}$ only had a few residues of interest that belong to the CBR3 loop (residues K267 and K269) and one of the ATP-binding sites (residues K66 and Y68). Of relevance is the residue Q97 that also appears in R130Q mutation pool with a 75% frequency. $PTEN^{R173C}$ also had a small number of residues within the active-site, WPD loop (residues Y88, F90, N94, and Q97), and residues in an ATP-binding site (N63, Y65, K66, and I67) as part of its pool. Of importance was the appearance of W274 and D92 in this pool with a frequency of 65% and 68%, respectively. Lastly, $PTEN^{R173H}$ comprised of several residues in the active-site appear in its frequency of nodes pool of paths.

Specifically, the CBR3 loop (residues K260, K267, K269, and F271), P loop (residues H123 and R130), WPD loop (residues Y88, F90, P95, and P96), and an ATP-binding site (residues Y68 and C71), with R130 and W274 appearing at a frequency of 45% and 58%. These residues all play a critical role in the salient communication pathway as outlined for each of the systems.

4.3.3 Community network analysis

A set of seven independent all-atom molecular dynamics (MD) simulations were conducted for wild-type PTEN (apo), PTEN-PIP_3 complex, and $PTEN^{R130G}$, $PTEN^{R130L}$, $PTEN^{R130Q}$, $PTEN^{R173C}$ and $PTEN^{R173H}$ mutants. During the WT PTEN (apo) simulation, a rather large conformational change occurred within the active-site involving the P loop (residues 123-130), WPD loop (residues 88-98) and TI loops (residues 160-171) forming a more "open" active-site. A closer inspection revealed the destruction of a critical salt-bridge involving residues R130 and D92. The positioning of these two residues is essential for their interaction with the PIP_3 substrate and the overall function of PTEN. Further investigation of each of the remaining systems, revealed considerable changes in the active-site loop conformations. Moreover, our recent structure and global dynamics studies revealed inter-domain communication caused by mutations at positions 130 and 173 that lead to a dramatic effect on PTEN long-range communication within the protein. This motived a more thorough understanding of communication pathways to further elucidate the functional effect each mutation has in the development of endometriosis, endometrial and ovarian cancers.

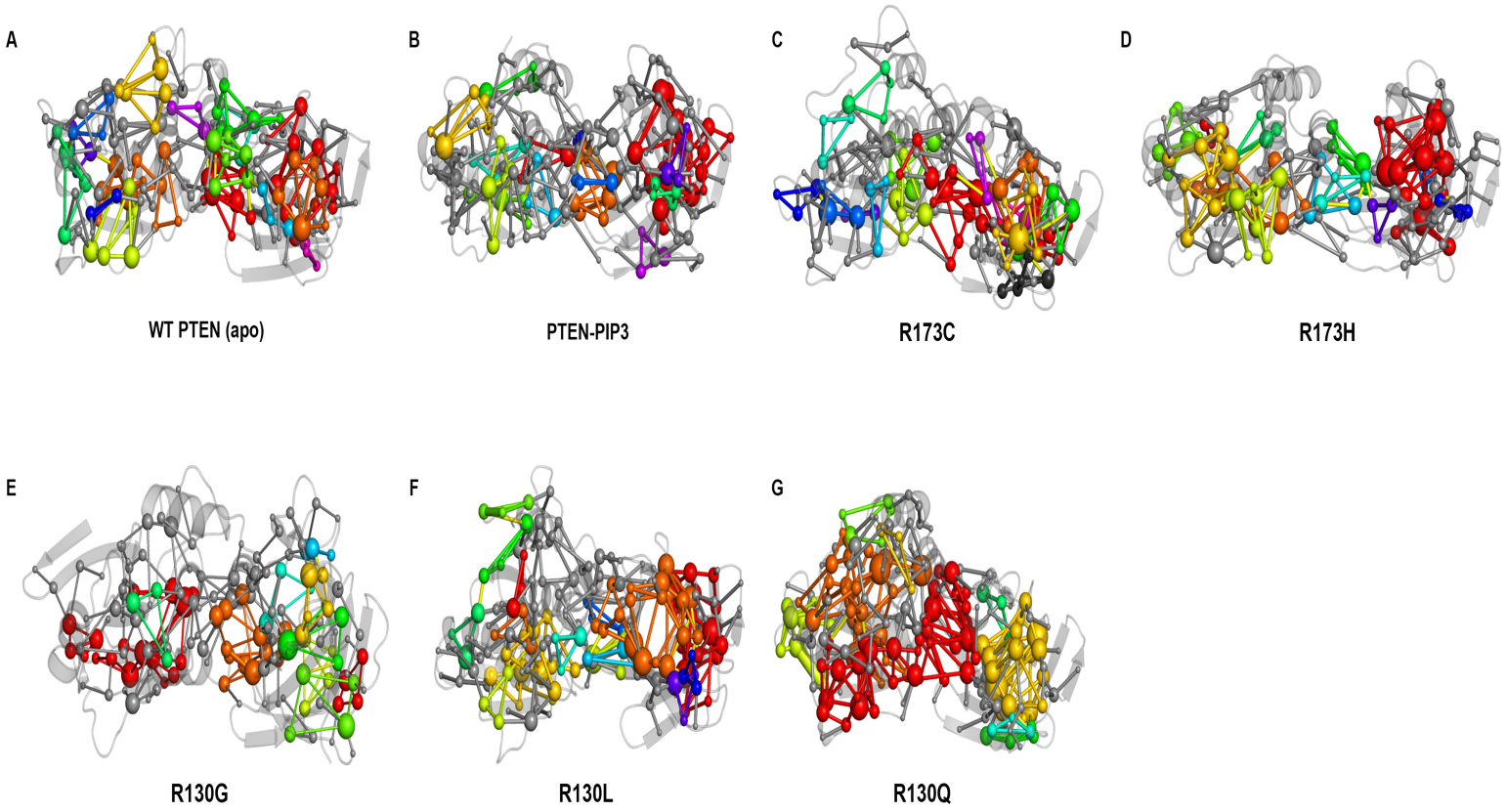
Figure 4.10— Conformational analysis of active-site loops



4.3.3.1 Community analysis of WT PTEN

A community network analysis approach[236,237,323,324] was applied to identify regions ('communities') within the protein that are closely correlate within the selected systems: WT PTEN (apo), PTEN-PIP_3 complex and $PTEN^{R130G}$, $PTEN^{R130L}$, $PTEN^{R130Q}$, $PTEN^{R173C}$, and $PTEN^{R173H}$ mutants. These communities are correlated with each other through key residues that establish contacts for long-range allosteric mechanisms.[328] Representative communication network analyses obtained from the simulations are shown in **Figure 4.11**. Communities are sets of highly interconnected vertices such that nodes belonging to the same community are densely linked to each other and poorly connected to nodes outside the community.[324] The protein network is divided into communities based on the flow of allosteric information that passes through each pair of nodes ('edge). This is defined as the number of shortest pathways that pass through the edge and is measured by the edge betweenness parameter.

Figure 4.11 Community network analysis



Our results reveal that each of the mutants displayed dissociation between both halves of their domains, more specifically within their inter-domain regions where as WT PTEN (apo) and PTEN-PIP_3 behave more as individual communities. Both of their domains appear as distinct though strongly connected communities. This is consistent with PTEN's functional nature of substrate binding and catalysis. The community connections within the active-site and inter-domain regions of the WT (apo) and PTEN-PIP_3 complex, highlights the effectiveness of the substrate-binding contacts are which allows for communication between the two domains. As indicated by the sizeable connectivity of communities within the inter-domain region of the apo and PTEN-PIP_3 complex, further confirms that the rich hydrogen-bond network within the inter-domain interface allows information to pass through inter-domain interface.

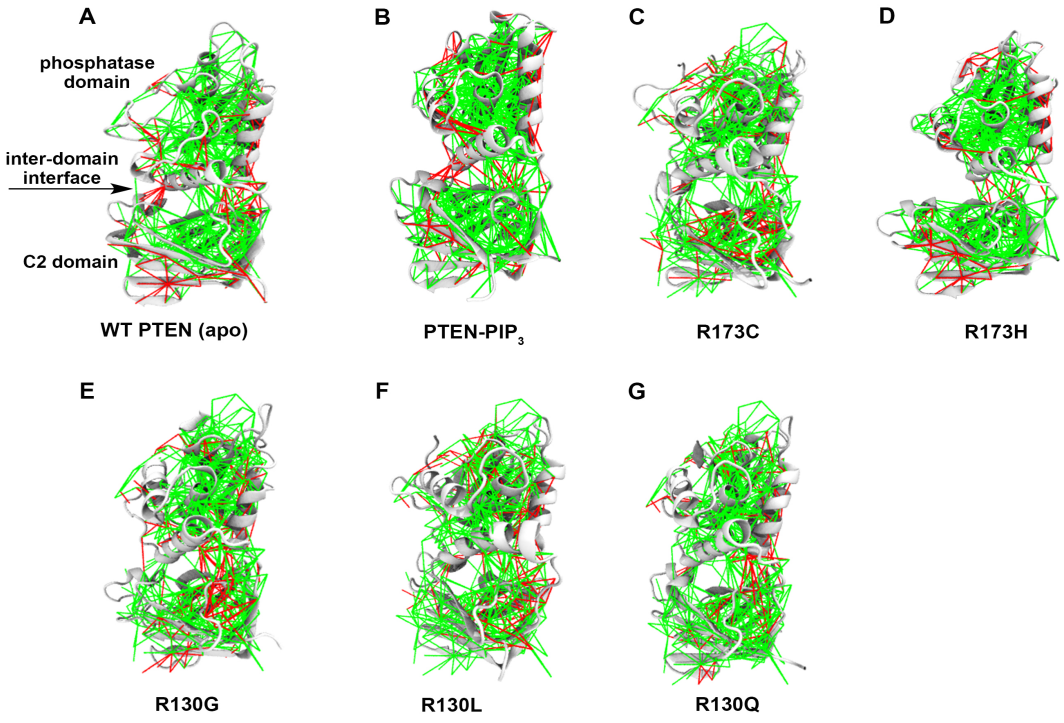
4.3.3.2 Community analysis of cancer-associated PTEN mutants

$PTEN^{R173C}$ had the most dissimilar communities possibly indicative of the loss of long-range communication pathway as a result of the mutation. Interestingly, $PTEN^{R130Q}$ exhibited a rather strongly connected community within the inter-domain interface. This is suggestive of these regions making a rather dense network possibly involved in the overall structural stability of the protein. The small number of communities in $PTEN^{R130G}$ as well as the lack of distinct communities and display of weak communications throughout the protein is most likely is due to the complete abolishment of charge as a result of the mutation. Charged or polar long-chain residues appear to be fundamental in transmitting allosteric information across different subunits. Similarly $PTEN^{R130L}$ and $PTEN^{R173H}$ had very few distinct communities except for some distinct communities within the CBR3 loop (residues 260-269) region indicating the significance of this region in transmitting information across the protein structure network.

4.3.4 Residual perturbation – local frustration analysis

To quantify the local frustration, we applied a frustratometer algorithm[325,326] to RMSD clusters of representative structures from each system. Residual Local Frustration is a measure of how 'frustrated' a residue is by verifying local/regional mutational or conformational changes shift residue energetics.[325] According to the principal of minimal frustration, which has its basis in the energy landscape theory, though minimally frustrated contacts (energetically favorable) are critical to stabilize the core of the protein, local clusters of highly frustrated contacts (energetically unfavorable) may have evolved facilitating protein motion and to allow the protein to modulate their function in response to their environment.[325] These highly frustrated contacts are thought to have profound effects on structural transitions and map to functional sites of the protein.

Figure 4.12 Residual frustration analysis



4.3.4.1 Residual frustration analysis of WT PTEN

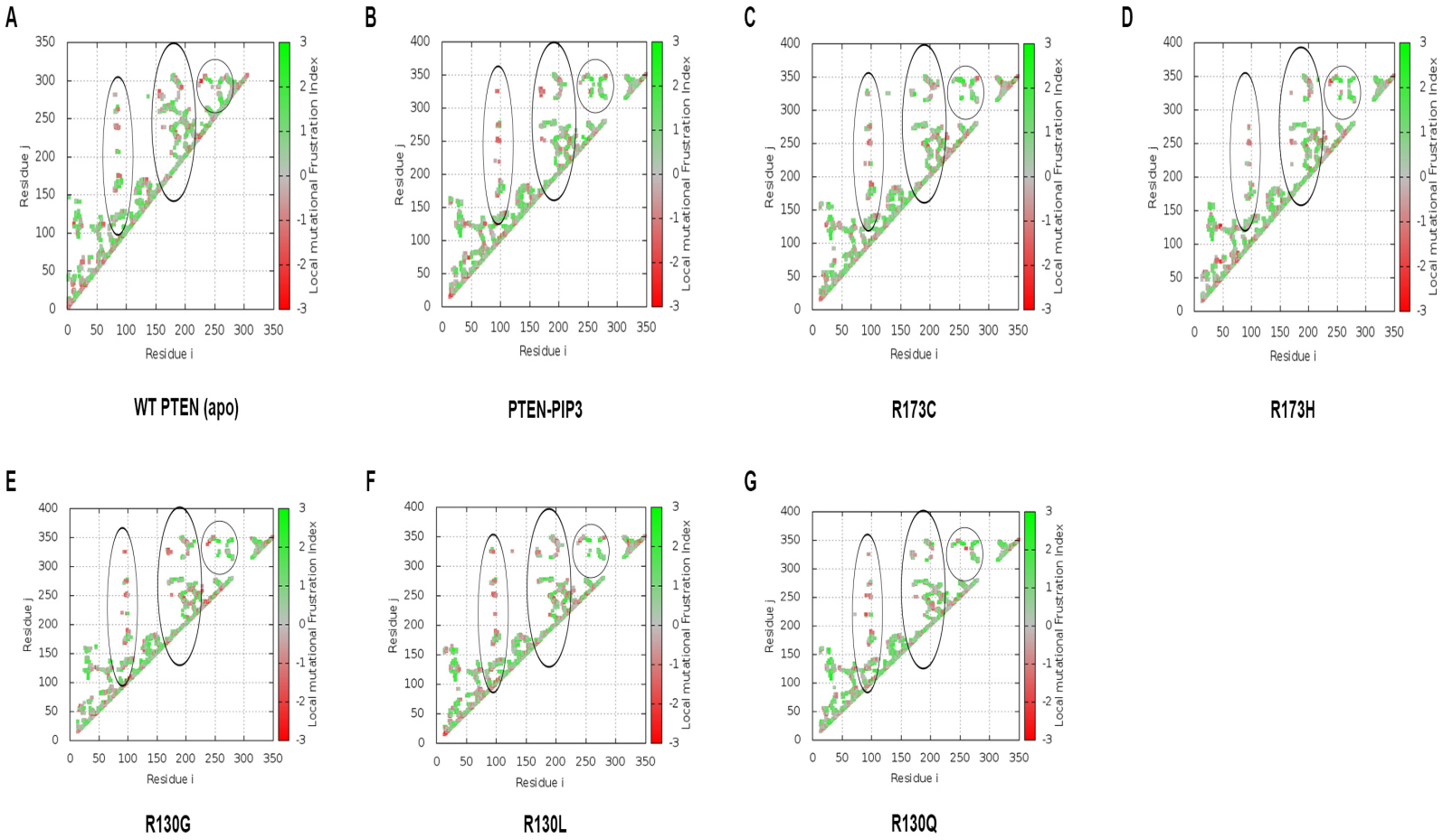
Figure 4.12 demonstrates the distribution of minimally (green) and highly (red) frustrated contacts within the active-site as well as the inter-domain interface of each of the investigated systems. The highly frustrated clusters most likely map to the active-site and inter-domain regions as a result of the interdependent conformational changes that must occur in protein-protein interactions or when interacting with the PIP_3 substrate. Since these regions are also conserved across multiple phosphatases, it highlights that regions of high-residual frustration may reflect an evolutionary requirement that often corresponds to functionally relevant sites.[325] Moreover, the conserved structural elements in the active-site, are coordinated by specific electrostatic interactions in order to achieve the precise positioning required for function. With

the domain-interface positioned adjacent to the active-site, its destabilization may be as a result of the increased stability of active-site substrate interactions as seen in **Figure 4.12B**.

4.3.4.2 Residual frustration analysis of cancer-associated PTEN mutants

Interestingly, the R130G/L/Q mutant systems encompass more frustrated contacts within the lower region of the adjoining inter-domain interface than other systems, more specifically the R173C/H mutations, which are located within the inter-domain interface. Moreover, the frustration index for each system as shown in **Figure 4.13**, illustrates three key areas that appear highly frustrated to varying degree in areas across the domains compared to the remainder of the protein. These regions encompass an area (residues 80-100) contiguous to the active-site, an ATP-binding motif, the inter-domain interface, and the CBR3 loop (residues 260-269). These results are in agreement with our previous studies where these mutated positions exhibited an increase in dynamics across the phosphatase-C2 domain interface further demonstrating that the integrity of this region is crucial to the overall structural stability of PTEN.

Figure 4.13 Frustration index



This index demonstrates how favorable a particular contact is relative to the set of all possible contacts in that location normalized utilizing the variance of that distribution.[325] A contact is defined as 'minimally frustrated' (green) if its native energy is at the lower end of distribution, having a frustration index of 0.78 or higher, indicating that other residues in that position would be unfavorable. Conversely, a contact is defined as 'highly frustrated' (red) if the native energy is at the other end of the distribution with a local frustration lower than -1, indicating that most other residues at that location would be favorable for folding than the native ones at standard deviation of distribution. If the native energy is between these two values, the contact is defined at 'neutral' (grey or not depicted). Interestingly, the PTEN-PIP_3 complex demonstrated a higher frustration index in the CBR3 loop region, illustrating that the binding of PIP_3 extends interdependent motions potentially corroborating a communication pathway essential to its catalytic function.

4.4 Conclusions

We utilized a combination of community network analysis, all-atom molecular dynamic simulations, and analysis of residual frustration to explore the dynamic ensemble of PTEN. Comparative analysis of WT PTEN (apo), PTEN-PIP_3 complex and $PTEN^{R130G}$, $PTEN^{R130L}$, $PTEN^{R130Q}$, $PTEN^{R173C}$, and $PTEN^{R173H}$ mutant systems identified difference in the long-range communication pathway as a result of ligand bound in active-site or mutation. Community analysis of PTEN mutants revealed community dissociation between both halves of their domains, more specifically within their inter-domain regions, whereas WT PTEN (apo) and PTEN in complex with PIP_3 revealed a more strongly connected community structure with convergence within the inter-domain interface (**Figure 4.11**).

The analysis of residual local frustration revealed that the R130G/L/Q mutant systems encompass more frustrated contacts within the lower region of the adjoining inter-domain interface than other systems, more specifically the R173C/H mutations, which are located within the inter-domain interface (**Figure 4.12**). Moreover, the frustration index illustrates three key areas that appear highly frustrated to varying degree in areas across the domains compared to the remainder of the protein (**Figure 4.13**). Interestingly, the PTEN-PIP_3 complex demonstrated a higher frustration index in the CBR3 loop region, illustrating that the binding of PIP_3 extends interdependent motions potentially corroborating a communication pathway essential to its catalytic function. However, in the case of the apo mutant systems, there exists a potential salient communication pathway underscoring a possible mutation-driven allosteric inter-domain interface. Interestingly, some flexible regions are separated by minimally frustrated contacts at the core of the protein; therefore it is crucial to investigate how information is transmitted along the structure through interdependent motions.

Our meta-path analysis revealed that though the WT PTEN (apo) has an identifiable communication pathway that is more spread out over the N-terminal region of the protein structure, it's strongest interaction is within the inter-domain interface (**Figure 4.7A**). The PTEN-PIP_3 complex has an even more distinguishable pathway that illustrates the strongest interaction to be in the WPD loop (residues 88-98) of the active-site that moves into the core of the protein and spreading throughout the C2 domain (**Figure 4.7B**). Notably, the meta-paths involving the PTEN mutants captured a rather significant scheme in their pathways of communication. Each of the mutants, had the active-site loops involved in their pathways and share several other features that consist of residues in their hubs containing both ATP-binding sites (residues 60-73 and 122-136) in addition to the CBR3 loop (residues 260-269). However, $PTEN^{R130Q}$ only had a few active-site hubs in its pathway (**Figure 4.7G**). The numbers of hubs in the WT PTEN (apo) and PTEN-PIP_3 complex systems were higher than the known catalytically inactive $PTEN^{R130G}$, which

had lowest number of hubs (**Table 4.2 and Figure 4.8**), this was consistent with the catalytically active mutants $PTEN^{R130L}$, $PTEN^{R130Q}$, $PTEN^{R173C}$, and $PTEN^{R173C}$.

These results further emphasize the role of the highly conserved R130 position both in structural stability, and intra-protein allosteric communication, deeming it as an integral hub residue participating in both inactive and active states. Interestingly, the semi-closed conformations represented in the $PTEN^{R130L}$, $PTEN^{R173C}$, and $PTEN^{R173C}$ mutants feature peculiar properties (a modest decrease in long-range paths) enforcing the notion that the global conformational changes between open and closed conformations is accompanied by internal alterations of the dynamical properties (**Figures 4.7 and 4.10**). Moreover, allosteric pathways typically involve highly frustrated clustered and specific polar amino acid residues such as R130 and R173. Overall our results demonstrate how all-atom MD, combined with PSN-ENM-NMA, and analysis of residual local frustration can be utilized to infer salient communication pathways in PTEN and suggest potential mechanisms for the coupling between the effects missense mutations have on global dynamics, structural communication, and function of PTEN. Future experiments might involve the characterization supporting the reported results that could potentially lead to new strategies in the development of allosteric drugs that target the inter-domain region of PTEN.

Chapter 5: Future directions

5.1 Chapter 3 future directions

Our study comprises an *in silico* effort in examining 13 somatic missense mutations in PTEN associated with endometriosis, endometrial cancer, and ovarian cancer. A combination of methods based on assessing protein structure stability, sequence evolutionary information, electrostatic potential, and protein global dynamics were employed to gain insight into the mutational effects that drive phenotypic outcome. Our observed predictions correlated with a decrease in protein structure stability in each of the associated mutations. I-Mutant 3.0 predicts all 13 mutations to be destabilizing. ENCoM predicts 6 mutations, (p.R130G/L/Q, p.R173C/H, and p.V191A) to be destabilizing. The effect of these mutations on structural stability and dynamics were further assessed indicating that the p.R130G/L/Q and p.R173C/H mutants have an increase in dynamics across the phosphatase-C2 domain interface.

Moreover, our results indicate that mutations at positions 130 and 173 may trigger certain types of allosteric changes as well as long-range communication effects that contribute to the loss of phosphatase activity toward PIP_3, thus giving rise to a severe phenotype. The results of our study can be exploited for future mutagenesis studies and as a framework to predict PTEN mutation severity beyond existing algorithms. Due to the potential significance of positions R130 and R173 being mutation-driven allosteric sites, the mechanism for their rescue of destabilized mutants, if not by global contribution to structure stability alone, poses a crucial question. Therefore, further analysis of the catalytic as well as tumor suppressor functions of mutations at positions 130 and 173 will aid in understanding the significance of mutational targeting of these positions in various cancers. A corresponding double mutant, at positions 130 and 173 will offer insight as to whether a rescue mutant confers potential contribution to structural stability ($\Delta\Delta G$). Of particular significance, is that the active-site also encompasses an ATP-binding motif

(residues 122-136) that modulates PTEN's subcellular localization. Cancer-associated mutations that lie within this ATP-binding motif, result in mutant PTEN that does not bind ATP resulting in altered signaling and growth; therefore underlying its critical pathogenic role in both heritable and sporadic carcinogenesis by PTEN nuclear mislocalization. In assessing the structural stability, change in charge, and global dynamics of PTEN ATP-binding motif mutations contained within this study such as p.R123Y, p.C124S, p. G129E/R, and p.R130G/L/Q mutations, represent an added advantage in understanding how the disruption of binding ATP results in endometriosis and cancer-associated phenotypes and present novel therapeutics in tumors with such PTEN alterations.

Additionally, our results provide a platform to further assess experimentally the effects of the mutations on stability of the inter-domain interface. To exploit this approach from an experimental standpoint, NMR could be utilized to determine the increase of flexibility of the inter-domain interface due to p.R173C and p.R173H mutations. Furthermore, shifts in pKa and hydrogen bond rearrangement at p.R130 mutations could also be tested by means of NMR along with thermal or denaturant unfolding methods.

5.2 Chapter 4 future directions

Our results encompass the combination of all-atom MD, combined with PSN-ENM-NMA, and analysis of residual local frustration to infer salient communication pathways in PTEN and suggest potential mechanisms for the coupling between the effects missense mutations have on global dynamics, structural communication, and function of PTEN. This combined approach allowed us to analyze the long-range communication within PTEN and associated mutants of the endometriosis, endometrial, and ovarian cancer phenotypes. Our choice of *in silico* generated mutants (R130G, R130L, R130Q, R173C, and R173H), derives from the high frequency

association of endometriosis and cancer at positions 130 and 173 (**Figure 4.1**) with each of the aforementioned phenotypes as well as our previous structural studies of PTEN.[300]

We identified distinct differences in the long-range communication pathway as a result of ligand bound in active-site (PTEN-PIP_3) or a mutation within the R130 and R173 positions. Comparative community analysis of PTEN mutants revealed community dissociation between both halves of their domains, more specifically within their inter-domain regions, whereas WT PTEN (apo) and PTEN in complex with PIP_3 revealed a more strongly connected community structure across the inter-domain interface (**Figure 4.11**). This demonstrates that the long-range communication is disrupted due to a mutational effect within the active-site or inter-domain region. Interestingly, the PTEN-PIP_3 complex demonstrated both a higher residual local frustration and frustration index in the CBR3 loop region, illustrating that the binding of PIP_3 extends interdependent motions potentially corroborating a communication pathway essential to its catalytic function. The meta-path analysis revealed that the WT PTEN (apo) had an identifiable communication with its strongest interaction within the inter-domain interface (**Figure 4.7A**). The PTEN-PIP_3 complex had an even more distinguishable pathway that illustrates the strongest interaction to be in the WPD loop (residues 88-98) that propagated throughout the C2 domain (**Figure 4.7B**). In particular, the PTEN mutants captured a rather significant scheme in their pathways of communication, which involved the active-site loops as a significant hub within their pathways.

Further analysis of the analyzed PTEN mutants coupled with either the PIP_3 substrate or ATP would provide further insight into the salient communication pathway for each of the aforementioned phenotypes as well as other cancer-associated mutations in the active-site and ATP-binding motif. Additionally, a comprehensive statistical coupling analysis on the co-evolution of R130 and R173 mutants could reveal potential new allosteric drug target sites in cancer therapy. These positions both reside in highly conserved regions of PTEN's structure.

Since a protein's function is dependent on it's three-dimensional structure as well as interaction of residues and their influence on structural stability, it is therefore essential to understand how compensatory mutations may occur to preserve or restore the function or structure of PTEN. In order to quantify evolutionary conserved network pathways that mediate allosteric communication and the co-variation between the two sites (R130 and R173), a mutual information (MI) model can be utilized.[329] There are many computational MI methods to exploit, however MI based on the incorporation and measurement of the residue background distribution as well as the measurement of individual residue physiochemical properties of two sites is critical to obtaining effective evaluation.[330] The R130 and R173 residues can in turn be exploited to further stabilize PTEN's structure and overall function in developing novel therapeutics in the treatment of cancer-associated PTEN mutants.

Chapter 6: References

1. Li J, Yen C, Liaw D, Podsypanina K, Bose S, Wang SI, Puc J, Miliaresis C, Rodgers L, McCombie R, Bigner SH, Giovanella BC, Ittmann M, Tycko B, Hibshoosh H, Wigler MH, Parsons R. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science 1997;275(5308):1943-1947.

2. Li J, Simpson L, Takahashi M, Miliaresis C, Myers MP, Tonks N, Parsons R. The PTEN/MMAC1 tumor suppressor induces cell death that is rescued by the AKT/protein kinase B oncogene. Cancer Res 1998;58(24):5667-5672.

3. Steck PA, Pershouse MA, Jasser SA, Yung WK, Lin H, Ligon AH, Langford LA, Baumgard ML, Hattier T, Davis T, Frye C, Hu R, Swedlund B, Teng DH, Tavtigian SV. Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. Nat Genet 1997;15(4):356-362.

4. Lee JO, Yang H, Georgescu MM, Di Cristofano A, Maehama T, Shi Y, Dixon JE, Pandolfi P, Pavletich NP. Crystal structure of the PTEN tumor suppressor: implications for its phosphoinositide phosphatase activity and membrane association. Cell 1999;99(3):323-334.

5. Gomperts BK, IM; Tatham PER. Signal Transduction. California: Elsevier Academic Press; 2002.

6. Rodriguez-Escudero I, Oliver MD, Andres-Pons A, Molina M, Cid VJ, Pulido R. A comprehensive functional analysis of PTEN mutations: implications in tumor- and autism-related syndromes. Hum Mol Genet 2011;20(21):4132-4142.

7. Georgescu MM. PTEN Tumor Suppressor Network in PI3K-Akt Pathway Control. Genes & cancer 2010;1(12):1170-1177.

8. Vazquez F, Ramaswamy S, Nakamura N, Sellers WR. Phosphorylation of the PTEN tail regulates protein stability and function. Mol Cell Biol 2000;20(14):5010-5018.

9. Lobo GP, Waite KA, Planchon SM, Romigh T, Houghton JA, Eng C. ATP modulates PTEN subcellular localization in multiple cancer cell lines. Hum Mol Genet 2008;17(18):2877-2885.

10. Lobo GP, Waite KA, Planchon SM, Romigh T, Nassif NT, Eng C. Germline and somatic cancer-associated mutations in the ATP-binding motifs of PTEN influence its subcellular localization and tumor suppressive function. Hum Mol Genet 2009;18(15):2851-2862.

11. He X, Ni Y, Wang Y, Romigh T, Eng C. Naturally occurring germline and tumor-associated mutations within the ATP-binding motifs of PTEN lead to oxidative damage of DNA associated with decreased nuclear p53. Hum Mol Genet 2011;20(1):80-89.

12. Gil A, Rodriguez-Escudero I, Stumpf M, Molina M, Cid VJ, Pulido R. A functional dissection of PTEN N-terminus: implications in PTEN subcellular targeting and tumor suppressor activity. PloS one 2015;10(4):e0119287.

13. Uversky VN, Dave V, Iakoucheva LM, Malaney P, Metallo SJ, Pathak RR, Joerger AC. Pathological unfoldomics of uncontrolled chaos: intrinsically disordered proteins and human diseases. Chemical reviews 2014;114(13):6844-6879.

14. Malaney P, Pathak RR, Xue B, Uversky VN, Dave V. Intrinsic disorder in PTEN and its interactome confers structural plasticity and functional versatility. Scientific reports 2013;3:2035.

15. Stuckey JA, Schubert HL, Fauman EB, Zhang ZY, Dixon JE, Saper MA. Crystal structure of Yersinia protein tyrosine phosphatase at 2.5 A and the complex with tungstate. Nature 1994;370(6490):571-575.

16. Yuvaniyama J, Denu JM, Dixon JE, Saper MA. Crystal structure of the dual specificity protein phosphatase VHR. Science 1996;272(5266):1328-1331.

17. Maehama T, Dixon JE. PTEN: a tumour suppressor that functions as a phospholipid phosphatase. Trends Cell Biol 1999;9(4):125-128.

18. Maehama T, Dixon JE. The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. J Biol Chem 1998;273(22):13375-13378.

19. Myers MP, Pass I, Batty IH, Van der Kaay J, Stolarov JP, Hemmings BA, Wigler MH, Downes CP, Tonks NK. The lipid phosphatase activity of PTEN is critical for its tumor supressor function. Proc Natl Acad Sci U S A 1998;95(23):13513-13518.

20. Zhang ZY. Mechanistic studies on protein tyrosine phosphatases. Progress in nucleic acid research and molecular biology 2003;73:171-220.

21. Tautz L, Critton DA, Grotegut S. Protein tyrosine phosphatases: structure, function, and implication in human disease. Methods Mol Biol 2013;1053:179-221.

22. Barford D, Flint AJ, Tonks NK. Crystal structure of human protein tyrosine phosphatase 1B. Science 1994;263(5152):1397-1404.

23. Denu JM, Dixon JE. Protein tyrosine phosphatases: mechanisms of catalysis and regulation. Current opinion in chemical biology 1998;2(5):633-641.

24. Fauman EB, Saper MA. Structure and function of the protein tyrosine phosphatases. Trends in biochemical sciences 1996;21(11):413-417.

25. Zhang ZY. Chemical and mechanistic approaches to the study of protein tyrosine phosphatases. Accounts of chemical research 2003;36(6):385-392.

26. Zhang ZY, Wang Y, Wu L, Fauman EB, Stuckey JA, Schubert HL, Saper MA, Dixon JE. The Cys(X)5Arg catalytic motif in phosphoester hydrolysis. Biochemistry 1994;33(51):15266-15270.

27. Chia JY, Gajewski JE, Xiao Y, Zhu HJ, Cheng HC. Unique biochemical properties of the protein tyrosine phosphatase activity of PTEN-demonstration of different active site

structural requirements for phosphopeptide and phospholipid phosphatase activities of PTEN. Biochim Biophys Acta 2010;1804(9):1785-1795.

28. Barford D, Das AK, Egloff MP. The structure and mechanism of protein phosphatases: insights into catalysis and regulation. Annual review of biophysics and biomolecular structure 1998;27:133-164.

29. Alicea-Velazquez NL, Boggon TJ. SHP family protein tyrosine phosphatases adopt canonical active-site conformations in the apo and phosphate-bound states. Protein and peptide letters 2013;20(9):1039-1048.

30. Koninckx PR, Meuleman C, Demeyere S, Lesaffre E, Cornillie FJ. Suggestive evidence that pelvic endometriosis is a progressive disease, whereas deeply infiltrating endometriosis is associated with pelvic pain. Fertil Steril 1991;55(4):759-765.

31. Eskenazi B, Warner ML. Epidemiology of endometriosis. Obstet Gynecol Clin North Am 1997;24(2):235-258.

32. Berkley KJ, Rapkin AJ, Papka RE. The pains of endometriosis. Science 2005;308(5728):1587-1589.

33. Giudice LC, Kao LC. Endometriosis. Lancet 2004;364(9447):1789-1799.

34. Bulun SE. Endometriosis. N Engl J Med 2009;360(3):268-279.

35. Nisolle M, Donnez J. Peritoneal endometriosis, ovarian endometriosis, and adenomyotic nodules of the rectovaginal septum are three different entities. Fertil Steril 1997;68(4):585-596.

36. Pritts EA, Taylor RN. An evidence-based evaluation of endometriosis-associated infertility. Endocrinology and metabolism clinics of North America 2003;32(3):653-667.

37. Nezhat F, Nezhat C, Pennington E. Laparoscopic proctectomy for infiltrating endometriosis of the rectum. Fertil Steril 1992;57(5):1129-1132.

38. Nezhat C, Nezhat F, Green B. Laparoscopic treatment of obstructed ureter due to endometriosis by resection and ureteroureterostomy: a case report. The Journal of urology 1992;148(3):865-868.

39. Nezhat F, Nezhat C, Levy JS. Laparoscopic treatment of symptomatic diaphragmatic endometriosis: a case report. Fertil Steril 1992;58(3):614-616.

40. Nezhat CR, Nezhat FR. Laparoscopic segmental bladder resection for endometriosis: a report of two cases. Obstet Gynecol 1993;81(5 (Pt 2)):882-884.

41. Nezhat C, Nezhat F, Pennington E, Nezhat CH, Ambroze W. Laparoscopic disk excision and primary repair of the anterior rectal wall for the treatment of full-thickness bowel endometriosis. Surgical endoscopy 1994;8(6):682-685.

42. Nezhat C, Kazerooni T, Berker B, Lashay N, Fernandez S, Marziali M. Laparoscopic management of hepatic endometriosis: report of two cases and review of the literature. Journal of minimally invasive gynecology 2005;12(3):196-200.

43. Medina MG, Lebovic DI. Endometriosis-associated nerve fibers and pain. Acta Obstet Gynecol Scand 2009;88(9):968-975.

44. Thibodeau LL, Prioleau GR, Manuelidis EE, Merino MJ, Heafner MD. Cerebral endometriosis. Case report. Journal of neurosurgery 1987;66(4):609-610.

45. Ichida M, Gomi A, Hiranouchi N, Fujimoto K, Suzuki K, Yoshida M, Nokubi M, Masuzawa T. A case of cerebral endometriosis causing catamenial epilepsy. Neurology 1993;43(12):2708-2709.

46. Sarma D, Iyengar P, Marotta TR, terBrugge KG, Gentili F, Halliday W. Cerebellar endometriosis. AJR Am J Roentgenol 2004;182(6):1543-1546.

47. Adamson GDK, S.; Hummelshoj L. Creating solutions in endometriosis: global collaboration through the World Endometriosis Research Foundation. Journal of Endometriosis 2010;2(1):3-6.

48. Garai J, Molnar V, Varga T, Koppan M, Torok A, Bodis J. Endometriosis: harmful survival of an ectopic tissue. Frontiers in bioscience : a journal and virtual library 2006;11:595-619.

49. Halban J. Metastatic hysteroadenosis. Wien Klin Wochenschr 1924;37:1205-1206.

50. Sampson JA. Metastatic or Embolic Endometriosis, due to the Menstrual Dissemination of Endometrial Tissue into the Venous Circulation. Am J Pathol 1927;3(2):93-110 143.

51. Gazvani R, Templeton A. New considerations for the pathogenesis of endometriosis. Int J Gynaecol Obstet 2002;76(2):117-126.

52. Sampson JA. Heterotropic or misplaced endometrial tissue. Am J Obstet Gynecol 1925;10(5):649-664.

53. Sasson IE, Taylor HS. Stem cells and the pathogenesis of endometriosis. Ann N Y Acad Sci 2008;1127:106-115.

54. Koninckx PR, Kennedy SH, Barlow DH. Endometriotic disease: the role of peritoneal fluid. Hum Reprod Update 1998;4(5):741-751.

55. Figueira PG, Abrao MS, Krikun G, Taylor HS. Stem cells in endometrium and their role in the pathogenesis of endometriosis. Ann N Y Acad Sci 2011;1221:10-17.

56. Burney RO, Giudice LC. Pathogenesis and pathophysiology of endometriosis. Fertil Steril 2012;98(3):511-519.

57. Martin JD, Jr., Hauck AE. Endometriosis in the male. The American surgeon 1985;51(7):426-430.

58. Sourial S, Tempest N, Hapangama DK. Theories on the pathogenesis of endometriosis. International journal of reproductive medicine 2014;2014:179515.

59. Nap AW, Groothuis PG, Demir AY, Evers JL, Dunselman GA. Pathogenesis of endometriosis. Best Pract Res Clin Obstet Gynaecol 2004;18(2):233-244.

60. Simpson JL, Elias S, Malinak LR, Buttram VC, Jr. Heritable aspects of endometriosis. I. Genetic studies. Am J Obstet Gynecol 1980;137(3):327-331.

61. Kennedy S. The genetics of endometriosis. J Reprod Med 1998;43(3 Suppl):263-268.

62. Nouri K, Ott J, Krupitz B, Huber JC, Wenzl R. Family incidence of endometriosis in first-, second-, and third-degree relatives: case-control study. Reproductive biology and endocrinology : RB&E 2010;8:85.

63. Coxhead DT, E. J. Familial inheritance of endometriosis in a British population. A case control study. J Obstet Gynaecol 1993;13:42-44.

64. Moen MH. Endometriosis in monozygotic twins. Acta Obstet Gynecol Scand 1994;73(1):59-62.

65. Hadfield RM, Mardon HJ, Barlow DH, Kennedy SH. Endometriosis in monozygotic twins. Fertil Steril 1997;68(5):941-942.

66. Treloar SA, O'Connor DT, O'Connor VM, Martin NG. Genetic influences on endometriosis in an Australian twin sample. sueT@qimr.edu.au. Fertil Steril 1999;71(4):701-710.

67. Kennedy S, Mardon H, Barlow D. Familial endometriosis. J Assist Reprod Genet 1995;12(1):32-34.

68. Stefansson H, Geirsson RT, Steinthorsdottir V, Jonsson H, Manolescu A, Kong A, Ingadottir G, Gulcher J, Stefansson K. Genetic factors contribute to the risk of developing endometriosis. Hum Reprod 2002;17(3):555-559.

69. Kennedy S, Bennett S, Weeks DE. Affected sib-pair analysis in endometriosis. Hum Reprod Update 2001;7(4):411-418.

70. Kennedy S. Genetics of endometriosis: a review of the positional cloning approaches. Semin Reprod Med 2003;21(2):111-118.

71. Bischoff FZ, Simpson JL. Heritability and molecular genetic studies of endometriosis. Hum Reprod Update 2000;6(1):37-44.

72. Bischoff FZ, Marquez-Do D, Dang D, Carson SA, Buster JE, Simpson JL. NAT2 and GSTM1 DNA polymorphisms: increased GSTM1 (active) genotypes in endometriosis. Fertil Steril 2002;77(2):S17-S17.

73. Baxter SW, Thomas EJ, Campbell IG. GSTM1 null polymorphism and susceptibility to endometriosis and ovarian cancer. Carcinogenesis 2001;22(1):63-65.

74. Nakago S, Hadfield RM, Zondervan KT, Mardon H, Manek S, Weeks DE, Barlow D, Kennedy S. Association between endometriosis and N-acetyl transferase 2 polymorphisms in a UK population. Mol Hum Reprod 2001;7(11):1079-1083.

75. Zondervan KT, Cardon LR, Kennedy SH. What makes a good case-control study? Design issues for complex traits such as endometriosis. Hum Reprod 2002;17(6):1415-1423.

76. Treloar SA, Wicks J, Nyholt DR, Montgomery GW, Bahlo M, Smith V, Dawson G, Mackay IJ, Weeks DE, Bennett ST, Carey A, Ewen-White KR, Duffy DL, O'Connor D T, Barlow DH, Martin NG, Kennedy SH. Genomewide linkage study in 1,176 affected sister pair families identifies a significant susceptibility locus for endometriosis on chromosome 10q26. Am J Hum Genet 2005;77(3):365-376.

77. Campbell IG, Thomas EJ. Endometriosis: candidate genes. Hum Reprod Update 2001;7(1):15-20.

78. Bischoff FZ, Heard M, Simpson JL. Somatic DNA alterations in endometriosis: high frequency of chromosome 17 and p53 loss in late-stage endometriosis. Journal of reproductive immunology 2002;55(1-2):49-64.

79. Sainz de la Cuesta R, Izquierdo M, Canamero M, Granizo JJ, Manzarbeitia F. Increased prevalence of p53 overexpression from typical endometriosis to atypical endometriosis and ovarian cancer associated with endometriosis. Eur J Obstet Gynecol Reprod Biol 2004;113(1):87-93.

80. Augoulea A, Alexandrou A, Creatsa M, Vrachnis N, Lambrinoudaki I. Pathogenesis of endometriosis: the role of genetics, inflammation and oxidative stress. Archives of gynecology and obstetrics 2012;286(1):99-103.

81. Bulun SE, Zeitoun KM, Takayama K, Sasano H. Molecular basis for treating endometriosis with aromatase inhibitors. Hum Reprod Update 2000;6(5):413-418.

82. Noble LS, Takayama K, Zeitoun KM, Putman JM, Johns DA, Hinshelwood MM, Agarwal VR, Zhao Y, Carr BR, Bulun SE. Prostaglandin E2 stimulates aromatase expression in endometriosis-derived stromal cells. The Journal of clinical endocrinology and metabolism 1997;82(2):600-606.

83. Golden RJ, Noller KL, Titus-Ernstoff L, Kaufman RH, Mittendorf R, Stillman R, Reese EA. Environmental endocrine modulators and human health: an assessment of the biological evidence. Crit Rev Toxicol 1998;28(2):109-227.

84. Crain DA, Janssen SJ, Edwards TM, Heindel J, Ho SM, Hunt P, Iguchi T, Juul A, McLachlan JA, Schwartz J, Skakkebaek N, Soto AM, Swan S, Walker C, Woodruff TK, Woodruff TJ, Giudice LC, Guillette LJ, Jr. Female reproductive disorders: the roles of endocrine-disrupting compounds and developmental timing. Fertil Steril 2008;90(4):911-940.

85. Rier SE, Martin DC, Bowman RE, Dmowski WP, Becker JL. Endometriosis in rhesus monkeys (Macaca mulatta) following chronic exposure to 2,3,7,8-tetrachlorodibenzo-p-

dioxin. Fundamental and applied toxicology : official journal of the Society of Toxicology 1993;21(4):433-441.

86. Hadfield R, Mardon H, Barlow D, Kennedy S. Delay in the diagnosis of endometriosis: a survey of women from the USA and the UK. Hum Reprod 1996;11(4):878-880.

87. Husby GK, Haugen RS, Moen MH. Diagnostic delay in women with pain and endometriosis. Acta Obstet Gynecol Scand 2003;82(7):649-653.

88. Nnoaham KE, Hummelshoj L, Webster P, d'Hooghe T, de Cicco Nardone F, de Cicco Nardone C, Jenkinson C, Kennedy SH, Zondervan KT, World Endometriosis Research Foundation Global Study of Women's Health c. Impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries. Fertil Steril 2011;96(2):366-373 e368.

89. Hudelist G, Fritzer N, Thomas A, Niehues C, Oppelt P, Haas D, Tammaa A, Salzer H. Diagnostic delay for endometriosis in Austria and Germany: causes and possible consequences. Hum Reprod 2012;27(12):3412-3416.

90. Rogers PA, D'Hooghe TM, Fazleabas A, Gargett CE, Giudice LC, Montgomery GW, Rombauts L, Salamonsen LA, Zondervan KT. Priorities for endometriosis research: recommendations from an international consensus workshop. Reproductive sciences 2009;16(4):335-346.

91. Sinaii N, Cleary SD, Ballweg ML, Nieman LK, Stratton P. High rates of autoimmune and endocrine disorders, fibromyalgia, chronic fatigue syndrome and atopic diseases among women with endometriosis: a survey analysis. Hum Reprod 2002;17(10):2715-2724.

92. Togashi K, Nishimura K, Kimura I, Tsuda Y, Yamashita K, Shibata T, Nakano Y, Konishi J, Konishi I, Mori T. Endometrial cysts: diagnosis with MR imaging. Radiology 1991;180(1):73-78.

93. Cody RF, Jr., Ascher SM. Diagnostic value of radiological tests in chronic pelvic pain. Bailliere's best practice & research Clinical obstetrics & gynaecology 2000;14(3):433-466.

94. Athey PA, Diment DD. The spectrum of sonographic findings in endometriomas. Journal of ultrasound in medicine : official journal of the American Institute of Ultrasound in Medicine 1989;8(9):487-491.

95. Fried AM, Rhodes RA, Morehouse IR. Endometrioma: analysis and sonographic classification of 51 documented cases. Southern medical journal 1993;86(3):297-301.

96. Mais V, Guerriero S, Ajossa S, Angiolucci M, Paoletti AM, Melis GB. The efficiency of transvaginal ultrasonography in the diagnosis of endometrioma. Fertil Steril 1993;60(5):776-780.

97. Eskenazi B, Warner M, Bonsignore L, Olive D, Samuels S, Vercellini P. Validation study of nonsurgical diagnosis of endometriosis. Fertil Steril 2001;76(5):929-935.

98. Dessole S, Farina M, Rubattu G, Cosmi E, Ambrosini G, Nardelli GB. Sonovaginography is a new technique for assessing rectovaginal endometriosis. Fertil Steril 2003;79(4):1023-1027.

99. Okaro E, Condous G. Diagnostic and therapeutic capabilities of ultrasound in the management of pelvic pain. Curr Opin Obstet Gynecol 2005;17(6):611-617.

100. Gordts S, Campo R, Rombauts L, Brosens I. Transvaginal hydrolaparoscopy as an outpatient procedure for infertility investigation. Hum Reprod 1998;13(1):99-103.

101. Cheng YM, Wang ST, Chou CY. Serum CA-125 in Preoperative patients at high risk for endometriosis. Obstet Gynecol 2002;99(3):375-380.

102. Bedaiwy MA, Falcone T. Laboratory testing for endometriosis. Clin Chim Acta 2004;340(1-2):41-56.

103. Bedaiwy MA, Falcone T, Sharma RK, Goldberg JM, Attaran M, Nelson DR, Agarwal A. Prediction of endometriosis with serum and peritoneal fluid markers: a prospective controlled trial. Hum Reprod 2002;17(2):426-431.

104. ASRM. Revised American Society for Reproductive Medicine classification of endometriosis: 1996. Fertil Steril 1997;67(5):817-821.

105. Bahamondes L, Petta CA, Fernandes A, Monteiro I. Use of the levonorgestrel-releasing intrauterine system in women with endometriosis, chronic pelvic pain and dysmenorrhea. Contraception 2007;75(6 Suppl):S134-139.

106. Verma A, Konje JC. Successful treatment of refractory endometriosis-related chronic pelvic pain with aromatase inhibitors in premenopausal patients. Eur J Obstet Gynecol Reprod Biol 2009;143(2):112-115.

107. Seal SL, Kamilya G, Mukherji J, De A, Ghosh D, Majhi AK. Aromatase inhibitors in recurrent ovarian endometriomas: report of five cases with literature review. Fertil Steril 2011;95(1).

108. Bulun SE, Yang S, Fang Z, Gurates B, Tamura M, Zhou J, Sebastian S. Role of aromatase in endometrial disease. The Journal of steroid biochemistry and molecular biology 2001;79(1-5):19-25.

109. Cook AS, Rock JA. Surgical-Management of Endometriosis. Semin Reprod Endocr 1991;9(2):138-148.

110. Zullo F, Palomba S, Zupi E, Russo T, Morelli M, Cappiello F, Mastrantonio P. Effectiveness of presacral neurectomy in women with severe dysmenorrhea caused by endometriosis who were treated with laparoscopic conservative surgery: a 1-year prospective randomized double-blind controlled trial. Am J Obstet Gynecol 2003;189(1):5-10.

111. Lichten EM, Bombard J. Surgical treatment of primary dysmenorrhea with laparoscopic uterine nerve ablation. J Reprod Med 1987;32(1):37-41.

112. Namnoum AB, Hickman TN, Goodman SB, Gehlbach DL, Rock JA. Incidence of symptom recurrence after hysterectomy for endometriosis. Fertil Steril 1995;64(5):898-902.

113. Rizk B, Fischer AS, Lotfy HA, Turki R, Zahed HA, Malik R, Holliday CP, Glass A, Fishel H, Soliman MY, Herrera D. Recurrence of endometriosis after hysterectomy. Facts, views & vision in ObGyn 2014;6(4):219-227.

114. Ripps BA, Martin DC. Correlation of focal pelvic tenderness with implant dimension and stage of endometriosis. J Reprod Med 1992;37(7):620-624.

115. Gao X, Yeh YC, Outley J, Simon J, Botteman M, Spalding J. Health-related quality of life burden of women with endometriosis: a literature review. Current medical research and opinion 2006;22(9):1787-1797.

116. Simoens S, Dunselman G, Dirksen C, Hummelshoj L, Bokor A, Brandes I, Brodszky V, Canis M, Colombo GL, DeLeire T, Falcone T, Graham B, Halis G, Horne A, Kanj O, Kjer JJ, Kristensen J, Lebovic D, Mueller M, Vigano P, Wullschleger M, D'Hooghe T. The burden of endometriosis: costs and quality of life of women with endometriosis and treated in referral centres. Hum Reprod 2012;27(5):1292-1299.

117. De Graaff AA, D'Hooghe TM, Dunselman GA, Dirksen CD, Hummelshoj L, Consortium WE, Simoens S. The significant effect of endometriosis on physical, mental and social wellbeing: results from an international cross-sectional survey. Hum Reprod 2013;28(10):2677-2685.

118. D'Hooghe TD, C. D.,; Dunselman, G. A. J.; De Graaff, A., W.E.R.F. EndoCost Consortium, Simoens S. The costs of endometriosis: it's the economy, stupid. Fertil Steril 2012;98(3):S218-S219.

119. Gao X, Outley J, Botteman M, Spalding J, Simon JA, Pashos CL. Economic burden of endometriosis. Fertil Steril 2006;86(6):1561-1572.

120. Simoens S, Hummelshoj L, D'Hooghe T. Endometriosis: cost estimates and methodological perspective. Hum Reprod Update 2007;13(4):395-404.

121. Simoens S, Hummelshoj L, Dunselman G, Brandes I, Dirksen C, D'Hooghe T. Endometriosis cost assessment (the EndoCost study): a cost-of-illness study protocol. Gynecol Obstet Invest 2011;71(3):170-176.

122. Simoens S, Meuleman C, D'Hooghe T. Non-health-care costs associated with endometriosis. Hum Reprod 2011;26(9):2363-2367.

123. Sinaii N, Cleary SD, Younes N, Ballweg ML, Stratton P. Treatment utilization for endometriosis symptoms: a cross-sectional survey study of lifetime experience. Fertil Steril 2007;87(6):1277-1286.

124. Guo SW. Recurrence of endometriosis and its control. Hum Reprod Update 2009;15(4):441-461.

125. Melin A, Sparen P, Persson I, Bergqvist A. Endometriosis and the risk of cancer with special emphasis on ovarian cancer. Hum Reprod 2006;21(5):1237-1242.

126. Brinton LA, Gridley G, Persson I, Baron J, Bergqvist A. Cancer risk after a hospital discharge diagnosis of endometriosis. Am J Obstet Gynecol 1997;176(3):572-579.

127. Swiersz LM. Role of endometriosis in cancer and tumor development. Ann N Y Acad Sci 2002;955:281-292; discussion 293-285, 396-406.

128. Baldi A, Campioni M, Signorile PG. Endometriosis: pathogenesis, diagnosis, therapy and association with cancer (review). Oncol Rep 2008;19(4):843-846.

129. Vlahos NF, Kalampokas T, Fotiou S. Endometriosis and ovarian cancer: a review. Gynecol Endocrinol 2010;26(3):213-219.

130. Nezhat F, Apostol R, Mahmoud M, el Daouk M. Malignant transformation of endometriosis and its clinical significance. Fertil Steril 2014;102(2):342-344.

131. Sampson JA. Endometrial carcinoma of ovary arising in endometrial tissue in that organ. Am J Obstet Gynecol 1925;9(1):111-1141.

132. McCluggage WG, Desai V, Toner PG, Calvert CH. Clear cell adenocarcinoma of the colon arising in endometriosis: a rare variant of primary colonic adenocarcinoma. Journal of clinical pathology 2001;54(1):76-77.

133. Min KW, Koh YW, Ryu YJ, Hong SM, Kim KR, Sung CO. Primary clear cell adenocarcinoma arising from rectal endometriosis. Digestive endoscopy : official journal of the Japan Gastroenterological Endoscopy Society 2013;25(2):209-210.

134. Balat O, Kudelka AP, Edwards CL, Silva E, Kavanagh JJ. Malignant transformation in endometriosis of the urinary bladder: case report of clear cell adenocarcinoma. Eur J Gynaecol Oncol 1996;17(1):13-16.

135. Lah K, Desai D, Hadway P, Perry-Keene J, Coughlin G. Primary vesical clear cell adenocarcinoma arising in endometriosis: a rare case of mullerian origin. Anticancer Res 2013;33(2):615-617.

136. Kwon YS, Nam JH, Choi G. Clear cell adenocarcinoma arising in endometriosis of a previous episiotomy site. Obstet Gynecol 2008;112(2 Pt 2):475-477.

137. Shah C, Pizer E, Veljovich DS, Drescher CW, Peters WA, 3rd, Paley PJ. Clear cell adenocarcinoma of the vagina in a patient with vaginal endometriosis. Gynecol Oncol 2006;103(3):1130-1132.

138. Bolis GB, Maccio T. Clear cell adenocarcinoma of the vulva arising in endometriosis. A case report. Eur J Gynaecol Oncol 2000;21(4):416-417.

139. Alberto VO, Lynch M, Labbei FN, Jeffers M. Primary abdominal wall clear cell carcinoma arising in a Caesarean section scar endometriosis. Irish journal of medical science 2006;175(1):69-71.

140. Mert I, Semaan A, Kim S, Ali-Fehmi R, Morris RT. Clear cell carcinoma arising in the abdominal wall: two case reports and literature review. Am J Obstet Gynecol 2012;207(2):e7-9.

141. Stevens EE, Pradhan TS, Chak Y, Lee YC. Malignant transformation of endometriosis in a cesarean section abdominal wall scar: a case report. J Reprod Med 2013;58(5-6):264-266.

142. Berglund AS, Sparen P, Bergqvist A. Endometriosis and the risk of cancer. Hum Reprod 2003;18:80-80.

143. Gemmill JA, Stratton P, Cleary SD, Ballweg ML, Sinaii N. Cancers, infections, and endocrine diseases in women with endometriosis. Fertil Steril 2010;94(5):1627-1631.

144. Kokcu A. Relationship between endometriosis and cancer from current perspective. Archives of gynecology and obstetrics 2011;284(6):1473-1479.

145. Benoit L, Arnould L, Cheynel N, Diane B, Causeret S, Machado A, Collin F, Fraisse J, Cuisenier J. Malignant extraovarian endometriosis: a review. Eur J Surg Oncol 2006;32(1):6-11.

146. Fishman A, Demirel D, Laucirica R, Ramzy I, Klima T, Lyzak J, Kaplan AL. Malignant tumors arising in endometriosis: clinical-pathological study and flow cytometry analysis. Eur J Obstet Gynecol Reprod Biol 1996;70(1):69-74.

147. Obata K, Morland SJ, Watson RH, Hitchcock A, Chenevix-Trench G, Thomas EJ, Campbell IG. Frequent PTEN/MMAC mutations in endometrioid but not serous or mucinous epithelial ovarian tumors. Cancer Res 1998;58(10):2095-2097.

148. Somigliana E, Vigano P, Parazzini F, Stoppelli S, Giambattista E, Vercellini P. Association between endometriosis and cancer: a comprehensive review and a critical analysis of clinical and epidemiological evidence. Gynecol Oncol 2006;101(2):331-341.

149. Nezhat FR, Pejovic T, Reis FM, Guo SW. The Link Between Endometriosis and Ovarian Cancer Clinical Implications. International Journal of Gynecological Cancer 2014;24(4):623-628.

150. Leiserowitz GS, Gumbs JL, Oi R, Dalrymple JL, Smith LH, Ryu J, Scudder S, Russell AH. Endometriosis-related malignancies. Int J Gynecol Cancer 2003;13(4):466-471.

151. Mostoufizadeh M, Scully RE. Malignant tumors arising in endometriosis. Clin Obstet Gynecol 1980;23(3):951-963.

152. Sainz de la Cuesta R, Eichhorn JH, Rice LW, Fuller AF, Jr., Nikrui N, Goff BA. Histologic transformation of benign endometriosis to early epithelial ovarian cancer. Gynecol Oncol 1996;60(2):238-244.

153. Vercellini P, Parazzini F, Bolis G, Carinelli S, Dindelli M, Vendola N, Luchini L, Crosignani PG. Endometriosis and ovarian cancer. Am J Obstet Gynecol 1993;169(1):181-182.

154. Scott RB. Malignant changes in endometriosis. Obstet Gynecol 1953;2(3):283-289.

155. Stern RC, Dash R, Bentley RC, Snyder MJ, Haney AF, Robboy SJ. Malignancy in endometriosis: Frequency and comparison of ovarian and extraovarian types. International Journal of Gynecological Pathology 2001;20(2):133-139.

156. Dinulescu DM, Ince TA, Quade BJ, Shafer SA, Crowley D, Jacks T. Role of K-ras and Pten in the development of mouse models of endometriosis and endometrioid ovarian cancer. Nat Med 2005;11(1):63-70.

157. Yu HC, Lin CY, Chang WC, Shen BJ, Chang WP, Chuang CM, Task Force on Carcinogenesis of Endometrial C. Increased association between endometriosis and endometrial cancer: a nationwide population-based retrospective cohort study. Int J Gynecol Cancer 2015;25(3):447-452.

158. Wei JJ, Paintal A, Keh P. Histologic and immunohistochemical analyses of endometrial carcinomas: experiences from endometrial biopsies in 358 consultation cases. Archives of pathology & laboratory medicine 2013;137(11):1574-1583.

159. Prat J, Gallardo A, Cuatrecasas M, Catasus L. Endometrial carcinoma: pathology and genetics. Pathology 2007;39(1):72-87.

160. Lax SF. Molecular genetic pathways in various types of endometrial carcinoma: from a phenotypical to a molecular-based classification. Virchows Archiv : an international journal of pathology 2004;444(3):213-223.

161. Zheng W, Xiang L, Fadare O, Kong B. A proposed model for endometrial serous carcinogenesis. The American journal of surgical pathology 2011;35(1):e1-e14.

162. Zaino R, Whitney C, Brady MF, DeGeest K, Burger RA, Buller RE. Simultaneously detected endometrial and ovarian carcinomas--a prospective clinicopathologic study of 74 cases: a gynecologic oncology group study. Gynecol Oncol 2001;83(2):355-362.

163. Falkenberry SS, Steinhoff MM, Gordinier M, Rappoport S, Gajewski W, Granai CO. Synchronous endometrioid tumors of the ovary and endometrium. A clinicopathologic study of 22 cases. J Reprod Med 1996;41(10):713-718.

164. Grammatikakis I, Zervoudis S, Evangelinakis N, Tziortzioti V. Endometrium and ovarian cancer synchronous to endometriosis--a retrospective study of our experience of 7 years. J Med Life 2010;3(1):76-79.

165. Wei JJ, William J, Bulun S. Endometriosis and ovarian cancer: a review of clinical, pathologic, and molecular aspects. Int J Gynecol Pathol 2011;30(6):553-568.

166. Mabrouk M, Vicenzi C, Ferrini G, Geraci E, Forno SD, Caprara G, Montanri G, Seracchioli R. Mixed adenocarcinoma of the rectovaginal septum associated with endometriosis and endometrial carcinoma: a case report. Case reports in oncology 2011;4(1):149-154.

167. Kurman RJ, Kaminski PF, Norris HJ. The behavior of endometrial hyperplasia. A long-term study of "untreated" hyperplasia in 170 patients. Cancer 1985;56(2):403-412.

168. Sasaki H, Nishii H, Takahashi H, Tada A, Furusato M, Terashima Y, Siegal GP, Parker SL, Kohler MF, Berchuck A, et al. Mutation of the Ki-ras protooncogene in human endometrial hyperplasia and carcinoma. Cancer Res 1993;53(8):1906-1910.

169. Maxwell GL, Risinger JI, Gumbs C, Shaw H, Bentley RC, Barrett JC, Berchuck A, Futreal PA. Mutation of the PTEN tumor suppressor gene in endometrial hyperplasias. Cancer Res 1998;58(12):2500-2503.

170. Seidman JD. Prognostic importance of hyperplasia and atypia in endometriosis. Int J Gynecol Pathol 1996;15(1):1-9.

171. Kvaskoff M, Mu F, Terry KL, Harris HR, Poole EM, Farland L, Missmer SA. Endometriosis: a high-risk population for major chronic diseases? Hum Reprod Update 2015;21(4):500-516.

172. Tone AA, Salvador S, Finlayson SJ, Tinker AV, Kwon JS, Lee CH, Cohen T, Ehlen T, Lee M, Carey MS, Heywood M, Pike J, Hoskins PJ, Stuart GC, Swenerton KD, Huntsman DG, Gilks CB, Miller DM, McAlpine JN. The role of the fallopian tube in ovarian cancer. Clinical advances in hematology & oncology : H&O 2012;10(5):296-306.

173. Pearce CL, Templeman C, Rossing MA, Lee A, Near AM, Webb PM, Nagle CM, Doherty JA, Cushing-Haugen KL, Wicklund KG, Chang-Claude J, Hein R, Lurie G, Wilkens LR, Carney ME, Goodman MT, Moysich K, Kjaer SK, Hogdall E, Jensen A, Goode EL, Fridley BL, Larson MC, Schildkraut JM, Palmieri RT, Cramer DW, Terry KL, Vitonis AF, Titus LJ, Ziogas A, Brewster W, Anton-Culver H, Gentry-Maharaj A, Ramus SJ, Anderson AR, Brueggmann D, Fasching PA, Gayther SA, Huntsman DG, Menon U, Ness RB, Pike MC, Risch H, Wu AH, Berchuck A, Consortium OCA. Association between endometriosis and risk of histological subtypes of ovarian cancer: a pooled analysis of case-control studies. Lancet Oncol 2012;13(4):385-394.

174. Levanon K, Crum C, Drapkin R. New insights into the pathogenesis of serous ovarian cancer and its clinical impact. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2008;26(32):5284-5293.

175. Gadducci A, Lanfredini N, Tana R. Novel insights on the malignant transformation of endometriosis into ovarian carcinoma. Gynecol Endocrinol 2014;30(9):612-617.

176. Sato N, Tsunoda H, Nishida M, Morishita Y, Takimoto Y, Kubo T, Noguchi M. Loss of heterozygosity on 10q23.3 and mutation of the tumor suppressor gene PTEN in benign endometrial cyst of the ovary: possible sequence progression from benign endometrial cyst to endometrioid carcinoma and clear cell carcinoma of the ovary. Cancer Res 2000;60(24):7052-7056.

177. Ogawa S, Kaku T, Amada S, Kobayashi H, Hirakawa T, Ariyoshi K, Kamura T, Nakano H. Ovarian endometriosis associated with ovarian carcinoma: a clinicopathological and immunohistochemical study. Gynecol Oncol 2000;77(2):298-304.

178. Gil A, Andres-Pons A, Pulido R. Nuclear PTEN: a tale of many tails. Cell Death Differ 2007;14(3):395-399.

179. Myers MP, Stolarov JP, Eng C, Li J, Wang SI, Wigler MH, Parsons R, Tonks NK. P-TEN, the tumor suppressor from human chromosome 10q23, is a dual-specificity phosphatase. Proc Natl Acad Sci U S A 1997;94(17):9052-9057.

180. Wang X, Jiang X. PTEN: a default gate-keeping tumor suppressor with a versatile tail. Cell Res 2008;18(8):807-816.

181. Han SY, Kato H, Kato S, Suzuki T, Shibata H, Ishii S, Shiiba K, Matsuno S, Kanamaru R, Ishioka C. Functional evaluation of PTEN missense mutations using in vitro phosphoinositide phosphatase assay. Cancer Res 2000;60(12):3147-3151.

182. Risinger JI, Hayes AK, Berchuck A, Barrett JC. PTEN/MMAC1 mutations in endometrial cancers. Cancer Res 1997;57(21):4736-4738.

183. Kanamori Y, Kigawa J, Itamochi H, Shimada M, Takahashi M, Kamazawa S, Sato S, Akeshima R, Terakawa N. Correlation between loss of PTEN expression and Akt phosphorylation in endometrial carcinoma. Clin Cancer Res 2001;7(4):892-895.

184. Kong D, Suzuki A, Zou TT, Sakurada A, Kemp LW, Wakatsuki S, Yokoyama T, Yamakawa H, Furukawa T, Sato M, Ohuchi N, Sato S, Yin J, Wang S, Abraham JM, Souza RF, Smolinski KN, Meltzer SJ, Horii A. PTEN1 is frequently mutated in primary endometrial carcinomas. Nat Genet 1997;17(2):143-144.

185. Tashiro H, Blazes MS, Wu R, Cho KR, Bose S, Wang SI, Li J, Parsons R, Ellenson LH. Mutations in PTEN are frequent in endometrial carcinoma but rare in other common gynecological malignancies. Cancer Res 1997;57(18):3935-3940.

186. Stambolic V, Suzuki A, de la Pompa JL, Brothers GM, Mirtsos C, Sasaki T, Ruland J, Penninger JM, Siderovski DP, Mak TW. Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN. Cell 1998;95(1):29-39.

187. Castiblanco GA, Pires NY, Wistuba OI, Riquelme SE, Andrade ML, Corvalan RA. [Pathogenic role of PTEN tumor suppressor gene in ovarian cancer associated to endometriosis]. Rev Med Chil 2006;134(3):271-278.

188. Mandai M, Yamaguchi K, Matsumura N, Baba T, Konishi I. Ovarian cancer in endometriosis: molecular biology, pathology, and clinical management. Int J Clin Oncol 2009;14(5):383-391.

189. Martini M, Ciccarone M, Garganese G, Maggiore C, Evangelista A, Rahimi S, Zannoni G, Vittori G, Larocca LM. Possible involvement of hMLH1, p16(INK4a) and PTEN in the malignant transformation of endometriosis. Int J Cancer 2002;102(4):398-406.

190. Zhang H, Zhao X, Liu S, Li J, Wen Z, Li M. 17betaE2 promotes cell proliferation in endometriosis by decreasing PTEN via NFkappaB-dependent pathway. Mol Cell Endocrinol 2010;317(1-2):31-43.

191. Montgomery GW, Nyholt DR, Zhao ZZ, Treloar SA, Painter JN, Missmer SA, Kennedy SH, Zondervan KT. The search for genes contributing to endometriosis risk. Hum Reprod Update 2008;14(5):447-457.

192. Nassiri IG, B.; Tavassoli, M. Bioinformatics Profiling of Missense Mutations. World Academy of Science, Engineering and Technology 2009;52:207-209.

193. Mah SKT, S.H. Computational Analysis of PTEN Gene Mutation. International Journal on Advanced Science Engineering Information Technology 2012;2(5):43-46.

194. Grimsley GR, Scholtz JM, Pace CN. A summary of the measured pK values of the ionizable groups in folded proteins. Protein Sci 2009;18(1):247-251.

195. Petsko GAaR, D. Protein Structure and Function: London: New Science Press; 2004.

196. Bashford D, Gerwert K. Electrostatic calculations of the pKa values of ionizable groups in bacteriorhodopsin. J Mol Biol 1992;224(2):473-486.

197. Antosiewicz J, McCammon JA, Gilson MK. The determinants of pKas in proteins. Biochemistry 1996;35(24):7819-7833.

198. Ullmann GMaB, E. Shifted Titration Curves in Proteins. In: Naray-Szabo G, editor. Protein Modelling. Springer International Publishing Switzaerland: Springer; 2014. p 152-153.

199. Laurents DV, Huyghues-Despointes BM, Bruix M, Thurlkill RL, Schell D, Newsom S, Grimsley GR, Shaw KL, Trevino S, Rico M, Briggs JM, Antosiewicz JM, Scholtz JM, Pace CN. Charge-charge interactions are key determinants of the pK values of ionizable groups in ribonuclease Sa (pI=3.5) and a basic variant (pI=10.2). J Mol Biol 2003;325(5):1077-1092.

200. Schutz CN, Warshel A. What are the dielectric "constants" of proteins and how to validate electrostatic models? Proteins-Structure Function and Bioinformatics 2001;44(4):400-417.

201. Shortle D. Mutational Studies of Protein Structures and Their Stabilities. Quarterly reviews of biophysics 1992;25(2):205-250.

202. Madura JDB, J.M.; Wade, R.C.; Davis, M.E.; Luty, B.A.; Ilin, A.; Antosiewicz, J.; Gilson, M.K.; Bagheri, B.; Scott, L.R., McCammon, J.A. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. Computer Physics Communications 1995;91(1-3):57-95.

203. Antosiewicz J, McCammon JA, Gilson MK. Prediction of pH-dependent properties of proteins. J Mol Biol 1994;238(3):415-436.

204. Antosiewicz J, Briggs JM, Elcock AH, Gilson MK, McCammon JA. Computing Ionization States of Proteins with a Detailed Charge Model. J Comput Chem 1996;17(14):1633-1644.

205. Deshpande N, Addess KJ, Bluhm WF, Merino-Ott JC, Townsend-Merino W, Zhang Q, Knezevich C, Xie L, Chen L, Feng Z, Green RK, Flippen-Anderson JL, Westbrook J, Berman HM, Bourne PE. The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. Nucleic Acids Res 2005;33(Database issue):D233-237.

206. Xiang Z. Advances in homology protein structure modeling. Current protein & peptide science 2006;7(3):217-227.

207. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14(1):33-38, 27-38.

208. Briggs JM. "Macromolecular Electrostatics." Lecture Notes for BCHS 4312.

209. Baker NA. Poisson-Boltzmann methods for biomolecular electrostatics. Methods Enzymol. Volume 383: Elsevier Science Publishing; 2004. p 94-118.

210. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22(22):4673-4680.

211. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 2009;25(9):1189-1191.

212. Hayward SadG, B. L. Normal Modes and Essential Dynamics. In: Kukol A, editor. Molecular Modeling of Proteins. Volume 443. Totowa, NJ: Humana Press; 2008. p 89-105.

213. Tirion MM. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. Phys Rev Lett 1996;77(9):1905-1908.

214. Bahar IaC, Q. Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems. Boca Raton, FL: Chapman & Hall/CRC; 2006. 406 p.

215. Haliloglu T, Bahar I, Erman B. Gaussian dynamics of folded proteins. Phys Rev Lett 1997;79(16):3090-3093.

216. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des 1997;2(3):173-181.

217. Eichinger BE. Elasticity Theory .3. Volume Exclusion and C2. Macromolecules 1972;5(5):647-+.

218. Flory PJ. Statistical Thermodynamics of Random Networks. Proceedings of the Rotal Society of London 1976;351:351-380.

219. Pearson DS. Scattered Intensity from a Chain in a Rubber Network. Macromolecules 1977;10(3):696-701.

220. Erman B, Kloczkowski A, Mark JE. Chain Dimensions and Fluctuations in Random Elastomeric Networks .2. Dependence of Chain Dimensions and Fluctuations on Macroscopic Strain. Macromolecules 1989;22(3):1432-1437.

221. Erman BaM, J.E. Structure and Properties of Rubberlike Networks. New York: Oxford University Press; 1997.

222. Godsil CaR, G. Algebraic Graph Theory. New York: Springer; 2001.

223. Eyal E, Yang LW, Bahar I. Anisotropic network model: systematic evaluation and a new web interface. Bioinformatics 2006;22(21):2619-2627.

224. Doruker P, Atilgan AR, Bahar I. Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: Application to alpha-amylase inhibitor. Proteins-Structure Function and Genetics 2000;40(3):512-524.

225. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophysical journal 2001;80(1):505-515.

226. Tama F, Sanejouand YH. Conformational change of proteins arising from normal mode calculations. Protein engineering 2001;14(1):1-6.

227. Xu C, Tobi D, Bahar I. Allosteric changes in protein structure computed by a simple mechanical model: hemoglobin T<-->R2 transition. J Mol Biol 2003;333(1):153-168.

228. Mouawad L, Perahia D. Motions in hemoglobin studied by normal mode analysis and energy minimization: evidence for the existence of tertiary T-like, quaternary R-like intermediate structures. J Mol Biol 1996;258(2):393-410.

229. Bahar I, Chennubhotla C, Erman B. Reply to 'Comment on elastic network models and proteins'. Phys Biol 2007;4(1):64-65.

230. del Sol A, Fujihashi H, Amoros D, Nussinov R. Residues crucial for maintaining short paths in network communication mediate signaling in proteins. Molecular systems biology 2006;2:2006 0019.

231. Vishveshwara S, Ghosh A, Hansia P. Intra and inter-molecular communications through protein structure network. Current protein & peptide science 2009;10(2):146-160.

232. Vendruscolo M, Dokholyan NV, Paci E, Karplus M. Small-world view of the amino acids that play a key role in protein folding. Physical review E, Statistical, nonlinear, and soft matter physics 2002;65(6 Pt 1):061910.

233. Amitai G, Shemesh A, Sitbon E, Shklar M, Netanely D, Venger I, Pietrokovski S. Network analysis of protein structures identifies functional residues. J Mol Biol 2004;344(4):1135-1146.

234. Brinda KV, Vishveshwara S. A network representation of protein structures: implications for protein stability. Biophysical journal 2005;89(6):4159-4170.

235. Kannan N, Vishveshwara S. Identification of side-chain clusters in protein structures by a graph spectral method. J Mol Biol 1999;292(2):441-464.

236. Seeber M, Felline A, Raimondi F, Mariani S, Fanelli F. WebPSN: a web server for high-throughput investigation of structural communication in biomacromolecules. Bioinformatics 2015;31(5):779-781.

237. Raimondi F, Felline A, Seeber M, Mariani S, Fanelli F. A Mixed Protein Structure Network and Elastic Network Model Approach to Predict the Structural Communication in Biomolecular Systems: The PDZ2 Domain from Tyrosine Phosphatase 1E As a Case Study. Journal of chemical theory and computation 2013;9(5):2504-2518.

238. Potapov V, Cohen M, Schreiber G. Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. Protein engineering, design & selection : PEDS 2009;22(9):553-560.

239. Serrano L, Day AG, Fersht AR. Step-wise mutation of barnase to binase. A procedure for engineering increased stability of proteins and an experimental analysis of the evolution of protein stability. J Mol Biol 1993;233(2):305-312.

240. Capriotti E, Fariselli P, Rossi I, Casadio R. A three-state prediction of single point mutations on protein stability changes. BMC Bioinformatics 2008;9 Suppl 2:S6.

241. Parthiban V, Gromiha MM, Schomburg D. CUPSAT: prediction of protein stability upon point mutations. Nucleic Acids Res 2006;34(Web Server issue):W239-242.

242. Frappier V, Najmanovich RJ. A coarse-grained elastic network atom contact model and its use in the simulation of protein dynamics and the prediction of the effect of mutations. PLoS computational biology 2014;10(4):e1003569.

243. Frappier V, Chartier M, Najmanovich RJ. ENCoM server: exploring protein conformational space and the effect of mutations on protein function and stability. Nucleic Acids Res 2015;43(W1):W395-400.

244. Allen MP. Introduction to Simulation Methods. In: Attig NB, K.; Grubmuller, H.; Kremer, K., editor; 2004; Gustav-Stresemann-Institut, Bonn Germany.

245. Hansson T, Oostenbrink C, van Gunsteren W. Molecular dynamics simulations. Curr Opin Struct Biol 2002;12(2):190-196.

246. Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC, Eastwood MP, Gagliardo J, Grossman JP, Ho CR, Ierardi DJ, Kolossvary I, Klepeis JL, Layman T, McLeavey C, Moraes MA, Mueller R, Priest EC, Shan YB, Spengler J, Theobald M, Towles B, Wang SC. Anton, a Special-Purpose Machine for Molecular Dynamics Simulation. Conf Proc Int Symp C 2007:1-12.

247. Adcock SA, McCammon JA. Molecular dynamics: survey of methods for simulating the activity of proteins. Chemical reviews 2006;106(5):1589-1615.

248. Frenkel DS, B. Understanding Molecular Simulation: From Algorithms to Application: San Diego: Academic Press; 1996.

249. Karplus M, Petsko GA. Molecular dynamics simulations in biology. Nature 1990;347(6294):631-639.

250. Leach AR. Molecular Modelling Principles and Applications; 2001.

251. Allen MPT, D.J. Computer Simulation of Liquids; 1987.

252. Verlet L. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. Phys Rev Lett 1967;159:98-103.

253. Hockney RW. The potential calculation and some applications. Methods Computational Physics 1970;9:136-211.

254. Vangunsteren WF, Berendsen HJC. Computer-Simulation of Molecular-Dynamics - Methodology, Applications, and Perspectives in Chemistry. Angew Chem Int Edit 1990;29(9):992-1023.

255. Darden T, York D, Pedersen L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. Journal of Chemical Physics 1993;98(12):10089-10092.

256. Toukmaji AY, Board JA. Ewald summation techniques in perspective: A survey. Computer Physics Communications 1996;95(2-3):73-92.

257. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem 2004;25(13):1656-1676.

258. Kuczera K. Molecular modeling of peptides. Methods Mol Biol 2015;1268:15-41.

259. Becker OMaK, M. A Guide to Biomolecular Simulations. Kaptein R, editor. Berlin: Springer; 2005.

260. Ryckaert JPC, C.; Berendsen, H.J.C. Numerical Integration of the Cartesian Eqyations of Motion of a System with constraints: Molecular Dynamics of n-alkanes. Journal of Computational Physics 1977;23:327-341.

261. Hess BB, H.; Berendsen, H.J.C.; Fraaije, J.G.E.M. LINCS: A linear constraint solver for molecular simulations. J Comput Chem 1997;18:1463-1472.

262. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. Journal of chemical theory and computation 2008;4(3):435-447.

263. Vega C, Abascal JLF. Simulating water with rigid non-polarizable models: a general perspective. Physical Chemistry Chemical Physics 2011;13(44):19663-19688.

264. Berendsen HJCP, J.P.M.; Van Gunsteren, W.F.; and Hermans, J. . Interaction Models for Water in Relation to Protein Hydration, "Intermolecular Forces". 1981:331-342.

265. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. Journal of Chemical Physics 1983;79(2):926-935.

266. Berensen HJCP, J.P.M.; Van Gunsteren, W.F.; Dinola, A.; Haak, J.R. Molecular Dynamics with Coupling to an External Bath. Journal of Chemical Physics 1984;81:3684-3690.

267. Phillips JL, Colvin ME, Newsam S. Validating clustering of molecular dynamics simulations using polymer models. BMC Bioinformatics 2011;12:445.

268. Daura XG, K. ; Jaun, B.; Seebach, D.; van Gunsteren, W. F. and Mark, A. E. . Peptide Folding: When Simulation Meets Experiment. Angew Chem Int Edit 1999;1/2(38):236-240.

269. Jarvis RAaP, E. A. Clustering using a similarity measure based on shared near neighbors. IEEE Transactions on Computers 1973;100(11):1025-1034.

270. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem 2005;26(16):1701-1718.

271. Shlens J. "A Tutorial on Principal Component Analysis". Center for Neural Science; 2003.

272. Hess B. Convergence of sampling in protein simulations. Physical review E, Statistical, nonlinear, and soft matter physics 2002;65(3 Pt 1):031910.

273. Jolliffe I. Principal Component Analysis. 2nd ed. New York, NY: Springer; 2002. p 30-55.

274. Wold SE, K.; Geladi, P. . Principal Componenet Analysis. Chemometrics and Intelligent Laboratory Systems 1987;2:37-52.

275. Papaleo E, Mereghetti P, Fantucci P, Grandori R, De Gioia L. Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: the myoglobin case. Journal of molecular graphics & modelling 2009;27(8):889-899.

276. Willner J, Wurz K, Allison KH, Galic V, Garcia RL, Goff BA, Swisher EM. Alternate molecular genetic pathways in ovarian carcinomas of common histological types. Human pathology 2007;38(4):607-613.

277. Lovly C, Sosman, J., Pao, W. My Cancer Genome, version 1.4.13.2487, http://www.mycancergenome.org 2015.

278. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, Teague JW, Campbell PJ, Stratton MR, Futreal PA. COSMIC: mining

complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res 2011;39(Database issue):D945-950.

279. Blachut-Okrasinska E, Lesyng B, Briggs JM, McCammon JA, Antosiewicz JM. Poisson-Boltzmann model studies of molecular electrostatic properties of the cAMP-dependent protein kinase. European biophysics journal : EBJ 1999;28(6):457-467.

280. Huyghues-Despointes BM, Thurlkill RL, Daily MD, Schell D, Briggs JM, Antosiewicz JM, Pace CN, Scholtz JM. pK values of histidine residues in ribonuclease Sa: effect of salt and net charge. J Mol Biol 2003;325(5):1093-1105.

281. Antosiewicz JB, J.M.; Elcock, A.H.; Gilson, M.K.; McCammon, J.A. Computing the Ionization States of proteins with a Detailed Charge Model. J Comput Chem 1996;17:1633-1644.

282. Brooks BRBREO, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy minimization, and dynamics calculations. J Comput Chem 1983;4:187-195.

283. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A 2001;98(18):10037-10041.

284. Magrane M, Consortium U. UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford) 2011;2011:bar009.

285. Pundir S, Magrane M, Martin MJ, O'Donovan C, UniProt C. Searching and Navigating UniProt Databases. Curr Protoc Bioinformatics 2015;50:1 27 21-10.

286. Ichiye T, Karplus M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Proteins 1991;11(3):205-217.

287. Bakan A, Meireles LM, Bahar I. ProDy: protein dynamics inferred from theory and experiments. Bioinformatics 2011;27(11):1575-1577.

288. Essmann UP, L.; Berkowitz, ML; Darden, T; Lee, H; Pedersen LG. A smooth particle mesh Ewald method. Journal of Chemical Physics 1995;103:8577-8593.

289. Maehama T, Taylor GS, Dixon JE. PTEN and myotubularin: novel phosphoinositide phosphatases. Annu Rev Biochem 2001;70:247-279.

290. Minaguchi T, Yoshikawa H, Oda K, Ishino T, Yasugi T, Onda T, Nakagawa S, Matsumoto K, Kawana K, Taketani Y. PTEN mutation located only outside exons 5, 6, and 7 is an independent predictor of favorable survival in endometrial carcinomas. Clin Cancer Res 2001;7(9):2636-2642.

291. Miller MP, Kumar S. Understanding human disease mutations through the use of interspecific genetic variation. Hum Mol Genet 2001;10(21):2319-2328.

292. Mooney SD, Klein TE. The functional importance of disease-associated mutation. BMC Bioinformatics 2002;3:24.

293. Iijima M, Huang YE, Luo HR, Vazquez F, Devreotes PN. Novel mechanism of PTEN regulation by its phosphatidylinositol 4,5-bisphosphate binding motif is critical for chemotaxis. J Biol Chem 2004;279(16):16606-16613.

294. Zhang XC, Piccini A, Myers MP, Van Aelst L, Tonks NK. Functional analysis of the protein phosphatase activity of PTEN. Biochem J 2012;444(3):457-464.

295. Waite KA, Eng C. Protean PTEN: form and function. Am J Hum Genet 2002;70(4):829-844.

296. Worby CA, Dixon JE. Pten. Annu Rev Biochem 2014;83:641-669.

297. Leslie NR, Longy M. Inherited PTEN mutations and the prediction of phenotype. Seminars in cell & developmental biology 2016;52:30-38.

298. McConechy MK, Ding J, Senz J, Yang W, Melnyk N, Tone AA, Prentice LM, Wiegand KC, McAlpine JN, Shah SP, Lee CH, Goodfellow PJ, Gilks CB, Huntsman DG. Ovarian and endometrial endometrioid carcinomas have distinct CTNNB1 and PTEN mutation profiles. Mod Pathol 2014;27(1):128-134.

299. Williams G. Elastic network model of allosteric regulation in protein kinase PDK1. BMC Struct Biol 2010;10:11.

300. Smith IN, Briggs JM. Structural mutation analysis of PTEN and its genotype-phenotype correlations in endometriosis and cancer. Proteins 2016;84(11):1625-1643.

301. Chi AS, Batchelor TT, Dias-Santagata D, Borger D, Stiles CD, Wang DL, Curry WT, Wen PY, Ligon KL, Ellisen L, Louis DN, Iafrate AJ. Prospective, high-throughput molecular profiling of human gliomas. Journal of neuro-oncology 2012;110(1):89-98.

302. Kato H, Kato S, Kumabe T, Sonoda Y, Yoshimoto T, Kato S, Han SY, Suzuki T, Shibata H, Kanamaru R, Ishioka C. Functional evaluation of p53 and PTEN gene mutations in gliomas. Clin Cancer Res 2000;6(10):3937-3943.

303. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, Olivi A, McLendon R, Rasheed BA, Keir S, Nikolskaya T, Nikolsky Y, Busam DA, Tekleab H, Diaz LA, Jr., Hartigan J, Smith DR, Strausberg RL, Marie SK, Shinjo SM, Yan H, Riggins GJ, Bigner DD, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE, Kinzler KW. An integrated genomic analysis of human glioblastoma multiforme. Science 2008;321(5897):1807-1812.

304. Wang SI, Puc J, Li J, Bruce JN, Cairns P, Sidransky D, Parsons R. Somatic mutations of PTEN in glioblastoma multiforme. Cancer Res 1997;57(19):4183-4186.

305. Zhang CL, Tada M, Kobayashi H, Nozaki M, Moriuchi T, Abe H. Detection of PTEN nonsense mutation and psiPTEN expression in central nervous system high-grade astrocytic tumors by a yeast-based stop codon assay. Oncogene 2000;19(38):4346-4353.

306. Hu X, Stebbins CE. Dynamics of the WPD loop of the Yersinia protein tyrosine phosphatase. Biophysical journal 2006;91(3):948-956.

307. Campbell RB, Liu F, Ross AH. Allosteric activation of PTEN phosphatase by phosphatidylinositol 4,5-bisphosphate. J Biol Chem 2003;278(36):33617-33620.

308. McConnachie G, Pass I, Walker SM, Downes CP. Interfacial kinetic analysis of the tumour suppressor phosphatase, PTEN: evidence for activation by anionic phospholipids. Biochem J 2003;371(Pt 3):947-955.

309. Vazquez F, Devreotes P. Regulation of PTEN function as a PIP3 gatekeeper through membrane interaction. Cell cycle 2006;5(14):1523-1527.

310. Redfern RE, Redfern D, Furgason ML, Munson M, Ross AH, Gericke A. PTEN phosphatase selectively binds phosphoinositides and undergoes structural changes. Biochemistry 2008;47(7):2162-2171.

311. Moncalero VL, Costanzo RV, Perandones C, Radrizzani M. Different conformations of phosphatase and tensin homolog, deleted on chromosome 10 (PTEN) protein within the nucleus and cytoplasm of neurons. PloS one 2011;6(4):e18857.

312. Nanda H, Heinrich F, Losche M. Membrane association of the PTEN tumor suppressor: Neutron scattering and MD simulations reveal the structure of protein-membrane complexes. Methods 2015;77-78:136-146.

313. Monod J, Wyman J, Changeux JP. On the Nature of Allosteric Transitions: A Plausible Model. J Mol Biol 1965;12:88-118.

314. Cooper A, Dryden DT. Allostery without conformational change. A plausible model. European biophysics journal : EBJ 1984;11(2):103-109.

315. Tsai CJ, del Sol A, Nussinov R. Allostery: absence of a change in shape does not imply that allostery is not at play. J Mol Biol 2008;378(1):1-11.

316. Popovych N, Sun S, Ebright RH, Kalodimos CG. Dynamically driven protein allostery. Nature structural & molecular biology 2006;13(9):831-838.

317. Gerstein M, Lesk AM, Chothia C. Structural mechanisms for domain movements in proteins. Biochemistry 1994;33(22):6739-6749.

318. Szilagyi A, Nussinov R, Csermely P. Allo-network drugs: extension of the allosteric drug concept to protein- protein interaction and signaling networks. Current topics in medicinal chemistry 2013;13(1):64-77.

319. Batey S, Randles LG, Steward A, Clarke J. Cooperative folding in a multi-domain protein. J Mol Biol 2005;349(5):1045-1059.

320. Nagradova NK. Interdomain interactions in oligomeric enzymes: creation of asymmetry in homo-oligomers and role in metabolite channeling between active centers of hetero-oligomers. FEBS Lett 2001;487(3):327-332.

321. Nussinov R, Tsai CJ. Unraveling structural mechanisms of allosteric drug action. Trends Pharmacol Sci 2014;35(5):256-264.

322. Csermely P, Korcsmaros T, Kiss HJ, London G, Nussinov R. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. Pharmacology & therapeutics 2013;138(3):333-408.

323. Raimondi F, Felline A, Fanelli F. Catching Functional Modes and Structural Communication in Dbl Family Rho Guanine Nucleotide Exchange Factors. Journal of chemical information and modeling 2015;55(9):1878-1893.

324. Fanelli F, Felline A, Raimondi F, Seeber M. Structure network analysis to gain insights into GPCR function. Biochem Soc Trans 2016;44(2):613-618.

325. Ferreiro DU, Hegler JA, Komives EA, Wolynes PG. Localizing frustration in native proteins and protein assemblies. Proc Natl Acad Sci U S A 2007;104(50):19819-19824.

326. Jenik M, Parra RG, Radusky LG, Turjanski A, Wolynes PG, Ferreiro DU. Protein frustratometer: a tool to localize energetic frustration in protein molecules. Nucleic Acids Res 2012;40(Web Server issue):W348-351.

327. Papa A, Wan L, Bonora M, Salmena L, Song MS, Hobbs RM, Lunardi A, Webster K, Ng C, Newton RH, Knoblauch N, Guarnerio J, Ito K, Turka LA, Beck AH, Pinton P, Bronson RT, Wei W, Pandolfi PP. Cancer-associated PTEN mutants act in a dominant-negative manner to suppress PTEN protein function. Cell 2014;157(3):595-610.

328. Sethi A, Eargle J, Black AA, Luthey-Schulten Z. Dynamical networks in tRNA:protein complexes. Proc Natl Acad Sci U S A 2009;106(16):6620-6625.

329. Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. Science 1999;286(5438):295-299.

330. Gao HY, Dou YC, Yang JL, Wang J. New methods to measure residues coevolution in proteins. BMC Bioinformatics 2011;12.